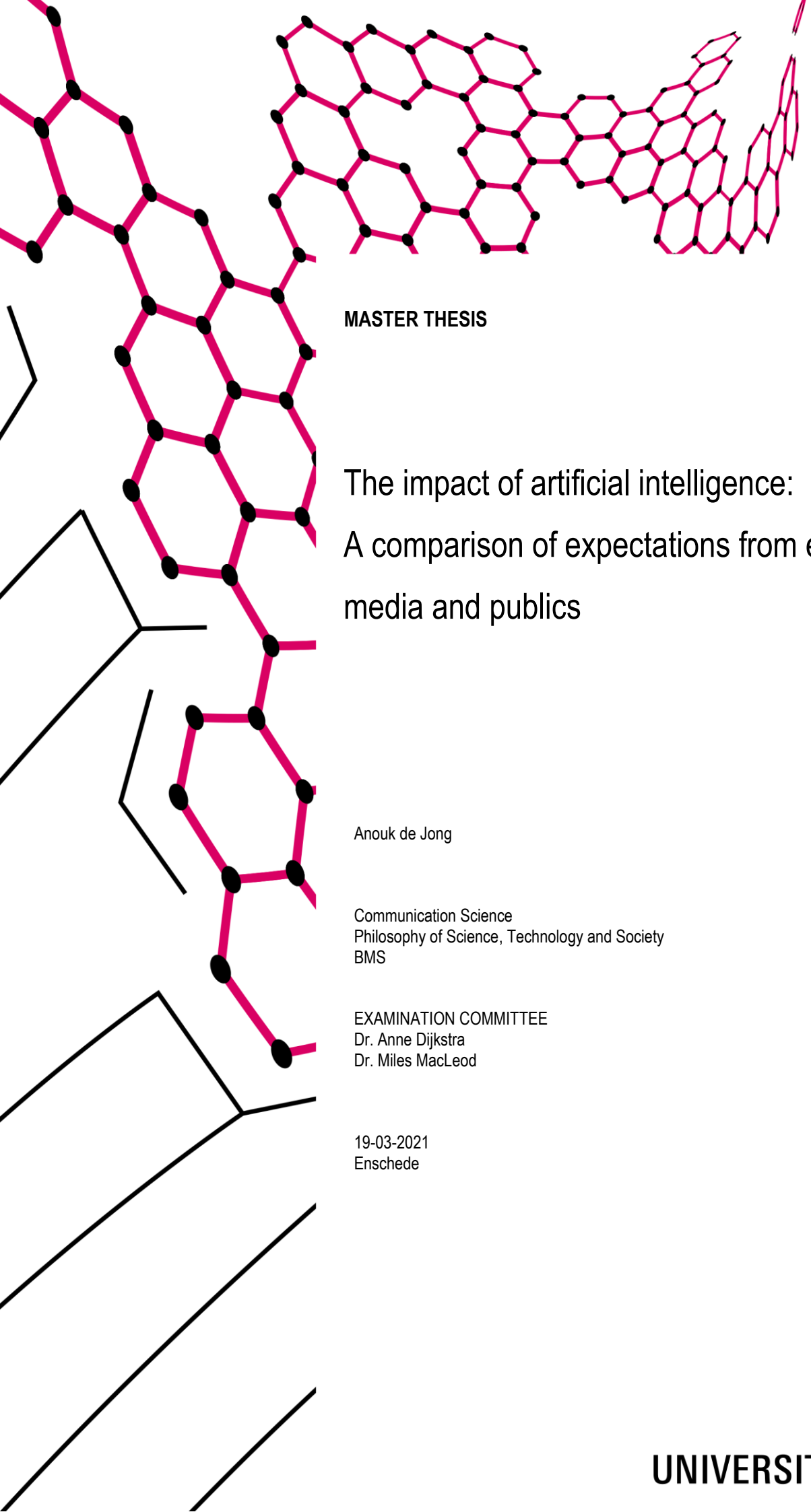


MASTER THESIS

The impact of artificial intelligence: A comparison of expectations from experts, media and publics

Anouk de Jong

Communication Science
Philosophy of Science, Technology and Society
BMS

EXAMINATION COMMITTEE
Dr. Anne Dijkstra
Dr. Miles MacLeod

19-03-2021
Enschede

UNIVERSITY OF TWENTE.

Abstract

The development and application of artificial intelligence (AI) has an increasing impact on society and on people's daily lives. News media play an important role in informing members of the public about new developments in AI and what impact these developments might have on their lives. The aim of this research was to study the role of communication and philosophy in increasing understanding of the science-society relationship. This was investigated by addressing two main research question. The first question was: How well aligned are philosophical discussions of AI with expert, media and public views and what consequences do current misalignments have for both philosophy and science-society relations? The second question was: How do views and expectations about AI discussed by experts, news media and publics relate to each other and what insight does this give for understanding the science-society relationship?

First of all, a literature analysis was conducted to define AI and to draw out the concepts of autonomy, responsibility, fairness, bias, explainability, and risk as main considerations in philosophical literature about AI. The quadruple helix was used as a representation of the science-society relationship. After the literature analysis, three empirical studies were conducted. The first study consisted of interviews with six experts in the academic, professional and governmental field of AI about their expectations for the development of AI and its societal impact. In the second study an in-depth media analysis (n53) was conducted about how Dutch newspaper articles portray AI and its impact. In the final study focus groups with Dutch citizens (n=18) were conducted to learn about their expectations of AI and its impact on society.

The results of these three studies showed that the six main concepts from philosophical literature reoccurred in the expert and public debates as well. Nevertheless there are some misalignments in how these concepts are discussed. The current misalignments can lead to negative impacts of AI being overlooked in the public debate and harm science-society relations. To prevent this, news media should add more nuance to their reports about the impact of AI and philosophical literature should focus more on weighing risks and benefits of applying AI in specific contexts, instead of focussing on what risks AI may pose in relation to abstract philosophical concepts.

From a communicative perspective, the comparison of the results showed that there is much overlap in the content discussed in news media and in the focus groups, pointing towards the reliance of laypeople on news media to receive information about AI. Furthermore, the focus on the philosophical concepts brought out nuances and depth in the analysis of the public debate about AI. This provides new insights about the science-society relationship that can be used to increase understanding of how to deal with emerging technologies in science communication.

Acknowledgements

Throughout, and even before, writing this thesis I have received a lot of help and support to make it possible to write one thesis to complete the master programmes of Communication Science and Philosophy of Science, Technology and Society.

First and foremost, I would like to thank my supervisors, Anne Dijkstra and Miles MacLeod, for your enthusiasm, encouragement and helpful feedback. I look forward to continue to work with you on my next research project.

In addition, I would like to thank everyone who helped me to make it possible to follow two master programmes at the same time and to combine everything I learned in one final thesis. This includes the study advisors, programme directors, examination boards and many helpful teachers and staff members at the University of Twente.

I would also like to thank everyone who participated in the interviews and focus groups for giving up your time and joining another online meeting in these times of working from home, in order to help me graduate.

Furthermore, I would like to thank all of my friends, for motivating me to continue, providing helpful tips and making studying a lot more fun.

Finally, I would like to thank my family and Wouter, for your continuous support and encouragement, for helping me through the harder times and for forcing me to take a break every now and then.

Index

1. Introduction

Due to the embedded nature of science and technology in various aspects of daily life, there is an increasing need for people to include scientific information when making important life decisions (National Academy of Sciences, 2017). Most people rely on science communication through media to receive information about science that is relevant for their life. However, the effectiveness of science communication depends on trust, including trust in the scientific source of information as well as trust in the medium of communication (Weingart & Guenther, 2016). Recently, this trust has been threatened by fundamental changes in how information is shared and an increase in the spread of misinformation about science (Scheufele & Krause, 2019). It is important to increase understanding of the science-society relationship in order to be able to face these challenges and communicate about science effectively.

One scientific topic that has recently received a lot of attention is the development of artificial intelligence (AI). AI is an emerging technology, that has an increasingly large impact on society. Applications of AI already influence various aspects of peoples' daily lives, including work, play, travel, communication, domestic tasks and security (Kitchin, 2017). Since its early stages of development, AI has been surrounded by speculations about what it could be and become (Natale & Ballatore, 2017). There has also been much attention to how AI might impact society and what ethical implications it might have. This makes it an interesting case to study from both a philosophical and communicative perspective.

The research problem that this thesis addresses concerns how information about artificial intelligence and its impact on society are discussed by philosophers, experts and in the public debate. The overarching research question is: *What insight does the case of AI give on the role of philosophy and communication in increasing understanding of the science-society relationship?* In order to investigate this, the thesis focuses on what expectation about AI and its impact are present in philosophical literature, among experts in the field of AI, in news media and among laypeople. The research problem has been divided into two main research questions, one relating to the research field of communication science and one relating to the domain of philosophy. In order to answer the research questions, sub-questions have been formulated for both main research question separately.

From a philosophical perspective the main research question is: *How well aligned are philosophical discussions of AI with expert, media and public views and what consequences do current misalignments have for both philosophy and the science-society relationship?* In order to answer this research question the following sub-questions will be answered: "What are the main considerations about the societal impact of AI in philosophical literature?", "What are the main considerations about the societal impact of AI among experts in the field?" and " What considerations about the societal impact of AI are apparent in the public debate about AI?".

From a communicative perspective, the main research question is: *How do views and expectations about AI discussed by experts, news media and publics relate to each other and what insight does this give for understanding the science-society relationship?* The following sub-questions will help to answer this question: “What views and expectations do experts in the field of AI have about artificial intelligence?”, “How do news media report about artificial intelligence?” and “What knowledge, views and expectations do laypeople have about artificial intelligence?”.

In order to address these research questions, several studies will be conducted and compared to each other. First of all, a literature analysis will be conducted to define artificial intelligence, bring out the most important concepts in philosophical literature about AI and provide an overview of existing literature on science communication about AI. Secondly, experts in the field of AI will be interviewed about their expectations for the development and societal impact of AI. Thirdly, a media analysis will be conducted to analyse how newspaper articles report about AI. Fourthly, focus groups will be conducted with Dutch citizens without expertise in AI, to learn about their expectations of AI and its impact. Finally, the results of these studies will be compared to each other, the theoretical and practical implications and limitations of this research will be discussed and suggestions for further research will be provided.

2. Theoretical Framework

2.1 Artificial intelligence

Artificial intelligence (AI) can be described as an umbrella term that is used to refer to any type of machine that is able to perform tasks that normally require human intelligence (Brennen et al., 2018; Helm et al., 2020). Such tasks include speech and image recognition, analysing large datasets and providing various recommendations (Helm et al., 2020, p. 69). However, there is no widespread agreement on the boundaries of what technologies can be classified as AI. What tasks normally require human intelligence is not self-evident and may change over time. In addition, there is no widespread consensus about a more comprehensive definition of artificial intelligence.

For the purpose of this research, the definition proposed by the European Commissions' High-Level Expert Group on Artificial Intelligence (AI HLEG) will be used. This definition is: "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" (AI HLEG, 2019a, p. 1). This definition provides a bit more guidance for what is seen as intelligent behaviour, although there is still room for debate about the degree of autonomy systems need to have. The AI HLEG (2019, p.1) also clarifies that AI systems van be purely applied in the virtual world or embedded in hardware devices, such as robots, drones or autonomous cars.

Within the field of AI, a distinction is often made between symbolic and non-symbolic AI (D'Souza, 2018). Symbolic AI is also called rule-based AI, since it works based on rules and facts that are put together in an algorithm by a person (D'Souza, 2018). For this type of algorithm people have to translate the relevant facts and rules into data the computer can understand and provide patterns, logical rules and calculations that the computer executes (D'Souza, 2018). Because of this, symbolic AI systems have trouble with dynamically changing facts and rules, it takes a long time to adapt the algorithm to new information (D'Souza, 2018).

Non-symbolic AI is often referred to as machine learning, because in this case raw data is provided which the computer uses to detect patterns and create its own representations (D'Souza, 2018). Because machine learning systems learn by themselves, it is easier for them to adapt to changing facts, rules and new conflicting data (D'Souza, 2018). However, these systems also require enormous amounts of data to work properly and the patterns and representations these systems create are often too abstract or complex for people to understand (D'Souza, 2018). It is also possible to combine symbolic with non-symbolic AI, by integrating representations that are understandable to people in machine learning algorithms (D'Souza, 2018)

2.2 Philosophical debate surrounding AI

The development of artificial intelligence (AI) has raised many philosophical questions. Within the public and professional discourse about AI there are some philosophical concepts that are central to the discussion. Multiple analyses have been made of which concepts are and should be considered in the development and implementation of AI. This chapter will provide an overview of the most important concepts that will be considered in this research.

The High-Level Expert Group on AI (AI HLEG) that was mentioned before, consists of experts from academia, industry and civil society appointed by the European Commission, to provide advice on the development and deployment of AI (AI HLEG, 2019b). This group selected four ethical principles based on relevant fundamental human rights that should be considered in the development and deployment of AI (AI HLEG, 2019b, p. 11). The ethical principles they selected are: respect for human autonomy, prevention of harm, fairness and explicability (AI HLEG, 2019b, p. 12). In addition, these principles have been translated into seven key requirements for AI systems, which are: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; accountability (AI HLEG, 2019b, p. 14).

Hayes, van de Poel and Steen (2020) provided a more extensive list of philosophical concepts related to AI that includes the principles that the AI HLEG selected. They investigated what values need to be taken into account when applying a value sensitive design approach to the application of machine learning algorithms in the domain of justice and security (Hayes et al., 2020, p. 1). They selected the values of accuracy, autonomy, privacy, fairness and equality, ownership and property, and accountability and transparency (Hayes et al., 2020, p. 2). Many of these values are relevant for the use of machine learning algorithms in other domains than that of justice and security, and the broader field of AI as well.

Vakkuri and Abrahamsson (2018) conducted a systematic mapping study to identify reoccurring keywords in 83 selected academic papers about ethics of AI. They used a list of 324 keywords that authors added to their articles in the databases and found that 37 of these keywords were used to describe multiple papers (Vakkuri & Abrahamsson, 2018, p. 4). The philosophical concepts that reoccurred most often were autonomy and responsibility, which were both used to describe five different papers (Vakkuri & Abrahamsson, 2018, p.4). The related concepts of consciousness, free will, existential risk, moral agency and moral patiency reoccurred in three papers (Vakkuri & Abrahamsson, 2018, p. 4). Since Vakkuri and Abrahamsson (2018) only analysed a relatively small amount of academic papers about the ethics of AI specifically, this does not provide a complete overview of the issues that are at stake in this case. Nevertheless, they provide a useful addition by distinguishing between autonomy and responsibility and emphasizing the importance of both concepts. This research will focus on the concepts of autonomy,

responsibility, fairness, bias, explainability and risk. These concepts were chosen based on a combination of the principles, values and keywords that were identified in the aforementioned analyses.

2.2.1 Autonomy

AI is regularly described as having autonomy, though it is often unclear what is meant by that (Johnson & Verdicchio, 2018, p. 639). In popular media as well as in scientific literature authors have expressed fears of AI becoming fully autonomous and making humans irrelevant (Johnson & Verdicchio, 2018, p. 639). There are even discussions about the possibility of AI becoming an existential threat by killing a large part of humanity (Vakkuri & Abrahamsson, 2018). When the term "artificial intelligence" was first introduced, it was expected that machines would be able to gain a type of intelligence that is similar to human intelligence (Helm et al., 2020, p. 69). The expectation was that one computer system would be able to outperform people in many different tasks. Instead of working towards such a general AI system, most research is currently focused on developing AI systems that can perform one specific task more quickly, efficiently or accurately than human experts (Helm et al., 2020, p. 70).

Even if AI does not become fully autonomous and out of control of humans, AI systems that are currently being deployed and developed may already influence the level of autonomy that people can exercise. The European High-Level Expert Group on Artificial Intelligence (AI HLEG 2019b) focused on this threat by including the principle of respect for human autonomy. In their explanation of this principle they argued that humans should be able to have full and effective self-determination and that they should be able to engage in the democratic process when interacting with AI systems (AI HLEG, 2019b, p. 12). They added that this means that AI systems should not unjustifiably manipulate, coerce, deceive or subordinate people (AI HLEG, 2019b, p. 12).

Hayes et al. (2020, p.7) defined autonomy as the ability for people to act intentionally and reflect consciously so they can live their life freely. They focussed specifically on decision-making algorithms in the judicial system and discussed how these algorithms may threaten the autonomy of both the decision maker and the person who is subject to the decision (Hayes et al., 2020). For decision makers there is a risk that they may automatically or uncritically trust the judgement of an algorithm above their own (Hayes et al., 2020, p. 7). In combination with the complexity and opacity of algorithms this may limit the autonomy of the decision maker, since they may not be able to critically reflect on the output of the algorithm (Hayes et al., 2020, p. 7).

For the subjects of decisions made by (or with the help of) machine learning algorithms there is a risk that their autonomy may be limited in different ways. In the domain of justice and security, algorithms can make subjects look suspicious, which diminishes the presumption of innocence and may foreclose future opportunities and freedoms for the subject (Hayes et al., 2020, p. 9). This foreclosing of future opportunities can be a risk of using algorithms in other situations, like the allocation of loans or the

selection of employees for job opportunities, as well. Johnson & Verdicchio (2018) argued that the widespread use of the concept 'autonomy' in relation to AI can cause confusion. They explained that the discussion about AI and autonomy is closely related to agency and responsibility (Johnson & Verdicchio, 2018, p. 639). They make a distinction between different types of agency that can provide clarity about responsibility when people interact with technology, autonomous or not (Johnson & Verdicchio, 2018, p. 640). This will be discussed in more detail in the next section on responsibility.

2.2.2 Responsibility

When AI applications are used it is often hard to figure out who is responsible if something goes wrong as a result of its use. When AI is seen as autonomous to a certain extent, this might lead to the conclusion that it is also at least partly responsible for its own actions. Johnson and Verdicchio (2018) distinguished between three different types of agency to clarify where the responsibility for AI applications lies. The first type of agency is causal agency, which means that someone or something plays a role in causing something to happen (Johnson & Verdicchio, 2018, p. 641). Causal agency can be attributed to any technology that influences if or how something happens (Johnson & Verdicchio, 2018, p. 641). The second type of agency is intentional agency, which adds the agent's intention as the beginning of the chain of causality that is also present in causal agency (Johnson & Verdicchio, 2018, p. 641). Since intentions are seen as mental states, intentional agency is usually only attributed to people (Johnson & Verdicchio, 2018, p. 641). Intentions are important, because in ethical and legal contexts, the type of intentions someone has determines whether they will be held responsible for causing something that happened (Johnson & Verdicchio, 2018, p. 641).

Johnson and Verdicchio (2018, p. 642) argued that technologies can play an important role in shaping people's intentions and making certain actions possible and that the concepts of causal and intentional agency do not suffice to accurately assign responsibility in such situations. To solve this problem they introduced a third type of agency called triadic agency (Johnson & Verdicchio, 2018, p. 642). Triadic agency assigns agency to the combination of a user, designer and artifact that caused something to happen together (Johnson & Verdicchio, 2018, p. 642). Only the humans in the triad (usually the user and/or designer) have intentional agency and can be assigned responsibility (Johnson & Verdicchio, 2018, p. 644). Johnson and Verdicchio (2018) also applied the concept of triadic agency to future scenarios in which the roles of user and designer might both be fulfilled by AI as well. They argued that in such cases responsibility should always be traced back to the human(s) who made the decision to design the AI in a certain way, since even a hypothetical super intelligent AI system cannot have intentional agency by itself and thus cannot be held morally and legally responsible (Johnson & Verdicchio, 2018, p. 645).

Hayes et al. (2020, p. 15) discussed responsibility in their examination of accountability and transparency. They defined accountability as a type of passive responsibility, meaning that agents can be held responsible and possibly be assigned blame

for something if they have moral agency, some causal relation to what happened and are suspected of some type of wrongdoing (Hayes et al., 2020, p. 15). Hayes et al. (2020, p.15) argued that information about an event or result and the people and things involved are needed in order to hold someone or a group of people accountable for the event or result. Following this, they argued that in situations that involve AI, this means that AI systems should be transparent (Hayes et al., 2020, p. 15). This will be discussed in more detail in the section about explainability.

2.2.3 Fairness

The concept of fairness is especially important in discussions about decision making algorithms. Saxena et al. (2020) compared three different definitions of fairness that have specifically been developed for decision making algorithms and conducted experiments to determine which definition people without expertise in AI preferred. The definitions they used focused on fairness as distributive justice, which prioritizes fair outcomes (Saxena et al., 2020, p. 2). The three definitions of fairness they compared are "treating similar individuals similarly", "never favor a worse individual over a better one" and "calibrated fairness" (Saxena et al., 2020, p. 3).

The first definition was proposed by Dwork et al. (2012) to develop algorithms that provide useful decisions that treat individuals with similar relevant characteristics in similar ways. The second definition was proposed by (Joseph et al., 2016) with the aim of making a fair algorithm that selects one candidate from a group (of people). They argued that a fair algorithm is one that always selects the candidate with the best relevant characteristics over the others (Joseph et al., 2016). Liu et al. (2017) based their definition of calibrated fairness on a combination of the previous two definitions. Calibrated fairness means that individuals are selected in proportion to their merit, so the best candidate receives the highest score and individuals with similar relevant characteristics get treated similarly. Saxena et al. (2020) found that the participants of their experiments preferred the calibrated fairness definition over the other two.

The discussion by Saxena et al. (2020) mainly concerns the public perception of definitions of fairness as they are currently used by computer scientists to create fair algorithms. More philosophically oriented discussions of fairness in relation to AI have been published as well. Binns (2018) studied fairness in machine learning from a political philosophy perspective. He explained that underlying patterns of discrimination in the world will likely be picked up as biases in machine learning processes and result in outputs that may lead to unfair treatment of certain groups and individuals (Binns, 2018, p. 1). Binns (2018, p. 9) further argued that current approaches to create fair machine learning risk focussing too much on narrow, static sets of protected classes based on law, without considering why these classes need special protection. He proposed that philosophical reflection on different theories of fairness and discrimination can help to address underlying issues in specific contexts (Binns, 2018, p. 9).

The AI HLEG (2019b) included the principle of fairness in their guidelines for trustworthy AI. They distinguished between substantive and procedural fairness, which should both be considered in the development of AI (AI HLEG, 2019b, p. 12). Substantive fairness entails that AI systems should ensure an equal and just distribution of benefits and costs, and ensure that there is no unfair bias, discrimination or stigmatization of individuals or groups (AI HLEG, 2019b, p. 12). Procedural fairness means that it is possible to contest and to effectively rectify decisions made by AI systems and the people using them (High-Level Expert Group on Artificial Intelligence, 2019b, p. 13). This requires that there is an identifiable entity that can be held accountable and that the decision-making process is explicable (High-Level Expert Group on Artificial Intelligence, 2019b, p. 13). The explicability or explainability of AI has implications for fairness as well as for autonomy, responsibility and the use of AI in general, therefore this will be discussed in detail as a separate concept.

Hayes et al. (2020, p.12) focused on fairness as an absence of discrimination or other types of arbitrary unequal treatment in their discussion of fairness and equality. They argued that people expect to be treated fairly in the sense that they are treated with equal regard, with the exception of situations that promote the interests of disadvantaged members of society (Hayes et al., 2020, p. 12). Hayes et al. (2020, p. 12) focus on what the AI HLEG (2019b) described as substantive fairness, arguing that AI systems might threaten fair treatment if they reproduce biases from their creators or training data. They further explained that discriminatory practices and limited perspectives can shape inaccurate machine learning models that disproportionally affect minorities and further increase unfair treatment of these groups (Hayes et al., 2020, p. 14).

2.2.4 Bias

In the discussion of fairness, bias was often mentioned as a possible cause of unfairness in AI, especially in the context of biases in decision making algorithms that lead to unfair results. Hayes et al. (2020) only discussed the concept of bias in relation to accuracy and fairness. In addition to their views on fairness, which were discussed in the previous section, they explained that algorithms might include biases because of design decisions, overrepresented or underrepresented data subjects or inaccurate data (Hayes et al., 2020, p. 4). They also emphasized the importance of the design of data abstractions and identified patterns, which can lead to the accidental inclusion of biases in algorithms (Hayes et al., 2020, p. 4). Since Hayes et al (2020) focused on decision making algorithms in the judicial systems, it is understandable that they emphasized how biases in algorithms can lead to unfair decisions. However, not all biases are unfair or harmful.

The ethics guidelines by the AI HLEG (2019b) mainly discussed bias as a cause of unfairness in AI as well, but in the glossary they explained that bias can be good or bad and intentional or unintentional. They also explained that bias does not necessarily relate to human bias or human-driven data collection, but can also arise through the contexts in which a system is used or through online learning and adaptation based on interaction (AI HLEG, 2019b, p. 36). Kitchin (2017, p.18) argued that algorithms should always be

understood as a relational and contingent element in the context in which they are developed and used. Since algorithms analyse and explore patterns in data, they categorize, sort and group data in certain ways, which includes certain biases (Kitchin, 2017, p. 18). Kitchin (2017, p. 19) concludes that algorithms may reform processes of sorting, classifying and differentiating data, but it is more likely that they deepen and accelerate these existing processes, which may be unfair.

Dobbe et al. (2018) provided a more in-depth explanation of different types of bias and how they might arise in machine learning algorithms. They argued that literature on fairness in AI has focused too much on how machine learning algorithms can inherit pre-existing biases from training data (Dobbe et al., 2018, p. 1). They stated that in addition to pre-existing biases, technical biases and emergent biases naturally occur in machine learning algorithms (Dobbe et al., 2018, p. 1). Dobbe et al. (2018, p. 2) explained that technical biases originate from the tools that AI developers use in the process of turning data into a model that can make decisions and predictions. They distinguished between four types of technical bias, namely, measurement bias, modelling bias, label bias and optimization bias (Dobbe et al., 2018, pp. 2–3). All of these technical biases arise in the development of machine learning algorithms. On the other hand, emergent biases only arise when machine learning algorithms are used in context (Dobbe et al., 2018, p. 3). As Dobbe et al. (2018, p. 3) explained machine learning systems act on their environment, but may also adapt based on feedback from that environment. Over time, this can lead to the formation of bias that could keep increasing over time as the feedback loop continues (Dobbe et al., 2018, p. 3).

2.2.5 Explainability

The explainability and transparency of AI is an important reoccurring topic in discussions about the fairness of AI. When deciding whether a decision made by an algorithm is fair, people usually want an explanation of how the algorithm arrived at this decision. In the case of machine learning algorithms this is complicated because these algorithms are often opaque. Regarding explainability, AI HLEG (2019b, p. 13) argued that the principle of explicability is essential for building and maintaining trust in AI systems. The principle of explicability includes that AI development processes need to be transparent, the capabilities and purposes of AI systems need to be openly communicated, and decisions made by AI systems need to be explainable to those affected by them as far as possible (AI HLEG, 2019b, p. 13).

As mentioned before, Hayes et al. (2020, p. 15) argued that transparency of algorithms and AI in general is necessary for accountability. In addition, they stated that transparency is important for many of the other values they discussed, including autonomy, fairness and privacy (Hayes et al., 2020, p. 16). Knowledge of an algorithm can help to counteract the ways in which algorithms may limit the autonomy of decision-makers and the subjects of decisions (Hayes et al., 2020, p. 16). In addition, it can help to judge if the decisions made by the algorithms are fair (Hayes et al., 2020, p. 16). Hayes et al. (2020, p.

15) use a definition of transparency as the possibility to get knowledge about some thing or event "characterized by availability, accessibility, understandability and explainability of relevant information".

AI systems, and especially machine learning algorithms, often complicate the process of getting relevant information. Burrell (2016, p. 1) focused on machine learning algorithms for classifications. She explained that these algorithms are usually opaque in the sense that recipients of a decision made by the algorithm do not know how or why the inputs of the algorithm lead to this decision (Burrell, 2016, p. 1). Burrell (2016, p. 1) distinguished between three different types of opacity that regularly occur in these algorithms and in AI in general. The first type is "opacity as intentional corporate or state secrecy" (Burrell, 2016, p.3). This type of opacity is present when the company or state that created the algorithm decides to keep the code secret, for example in order to have a competitive advantage, to prevent misuse or to hide secret intentions that the algorithm is used for (Burrell, 2016, p.4). The second type of opacity in algorithms is "opacity as technical illiteracy" (Burrell, 2016, p.4). This type of opacity is caused by the fact that very few people have the specialized skills and knowledge needed to create machine learning algorithms and to understand them properly (Burrell, 2016, p.4).

The final, most fundamental type of opacity is "opacity as the way algorithms operate at the scale of application". This type of opacity derives from how machine learning algorithms are created and how they work. Firstly, machine learning algorithms usually consist of many different components created by different people, which makes it very hard for one person to understand the complete system (Burrell, 2016, p.4). Secondly, machine learning algorithms that are useful need a very large amount of data, which interacts with the code used in the algorithm in complex ways (Burrell, 2016, p.5). Finally, Burrell (2016, p.5-7) argues that even if the code and the data of a machine learning algorithms are understandable separate from each other, the interplay between them is incomprehensible for people, because computers process information in a very different way.

2.2.6 Risk

The final concept in this research is risk. The principle of the prevention of harm that the High-Level Expert Group on Artificial Intelligence (AI HLEG, 2018) selected is included here, since the aim of this principle is to prevent the risk that AI might cause harm. The AI HLEG (2018, p. 12) report stated that AI systems should never cause or worsen harm, or negatively impact people in other ways. This means that AI systems should be developed and deployed in safe, secure and technically robust ways (AI HLEG, 2019b, p. 12). AI HLEG (2019b, p. 12) added that special attention should be paid to vulnerable persons and that other living beings and the natural environment should be considered as well. As this explanation shows, there is a risk that AI could cause harm in numerous areas and in various ways. Some risks have already been discussed in the sections on autonomy, responsibility,

fairness and explainability. However, there are some relevant risks AI could pose that fall outside of the scope of these concepts.

Firstly, there is a risk that AI could harm privacy. This risk has received much attention in ethics guidelines that have been developed for AI (Raab, 2020). Hayes et al (2020) also discussed privacy as one of the seven main values to take into account in the value sensitive design of AI. They explained that privacy includes ideas of control of and access to our physical space and personal information (Hayes et al., 2020, p. 10). They used privacy as contextual integrity of information, as proposed by Nissenbaum (2009), which means that privacy is respected if our personal information is transmitted by appropriate actors under appropriate principles, in a manner that adheres to the norms of the specific context we are in (Hayes et al., 2020, p. 10). Hayes et al. (2020, pp. 10-11) further discuss how the use of AI in the judicial system might threaten privacy as contextual integrity of information through the movement of personal data between contexts and the creation and categorization of groups. These risks may apply to applications of AI in other contexts as well.

The High-Level Expert Group on Artificial Intelligence (AI HLEG 2019b, p.10) discussed privacy and the right to a private life as part of the ethical principle of freedom of the individual. They also included privacy and data governance as one of their seven requirements of trustworthy AI (AI HLEG, 2019b, p. 14). This principle of privacy and data governance is closely related to the principle of the prevention of harm and includes respect for privacy, quality and integrity of data and access to data (AI HLEG, 2019b, p. 14). The AI HLEG (2019b, p. 14) argued that privacy and data protection should be guaranteed for information that the user initially provided, as well as for information AI systems may generate about the user over time through their interaction with the system.

Secondly, there is an environmental risk. The development and use of AI require computers and a lot of computing power. Ensmenger (2018) analysed the environmental history of computing by focusing on the material aspects of the use of computers and the internet. He noted that in 2003 the production of one desktop computer cost 240 kg of fossil fuels, 22 kg of chemicals and 1500 kg of water, excluding human labour (Ensmenger, 2018, p. 10). In addition, a lot of resources are needed for the storage and transmission of data via the internet. Ensmenger (2018, p. 4) reported that Googles data centres alone used more than 2.3 billion kw-h of electricity in 2011.

Strubell, Ganesh and McCallum (2020) researched the environmental cost of training machine learning algorithms by estimating the amount of energy required to train natural language processing (NLP) models. NLP models are types of machine learning algorithms that can recognize and make sense of written or spoken language. The accuracy of this type of algorithms increased drastically over the past few years due to the increase in available computing power (Strubell et al., 2020, p. 1). Strubell et al. (2020, p. 3) estimated that the most popular NLP models caused between 192 and 626,166 pounds of CO_2 emissions based on the hardware and the amount of power that was used and on the

training time of the algorithm. In comparison, on average a car causes 126,000 pounds of CO2 emissions in one lifetime, including fuel (Strubell et al., 2020, p. 1).

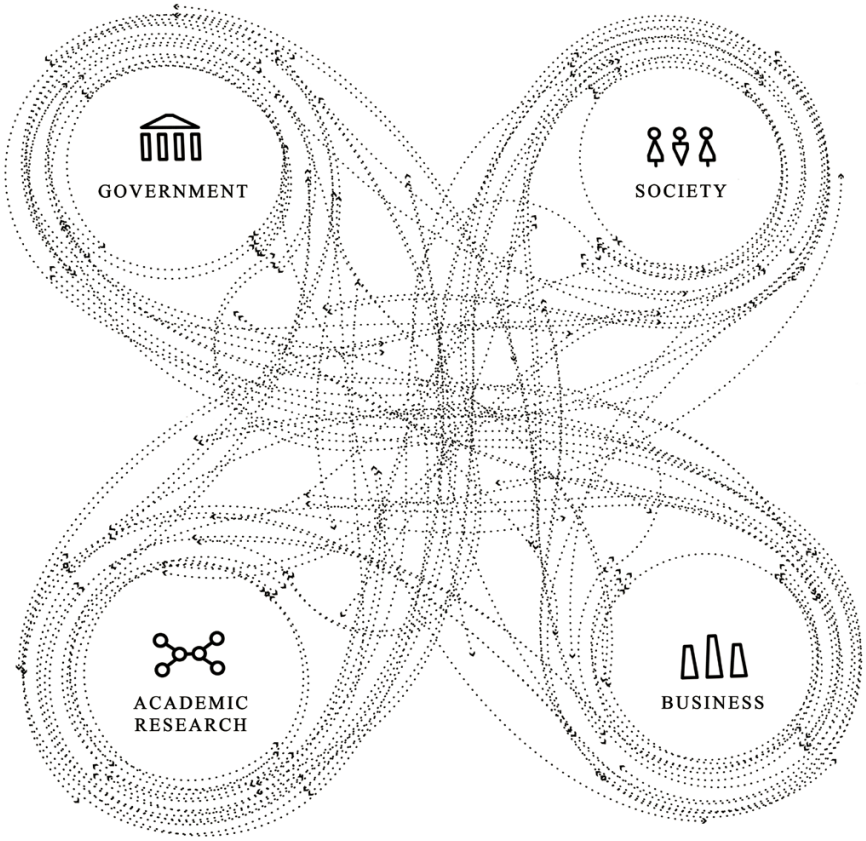
Finally, there are some other risks that are regularly mentioned in philosophical debates about the impact of AI that will not be discussed in detail. Risks related to the technical robustness, safety and accuracy of AI applications are often discussed as an important criterium for the use and development of AI (AI HLEG, 2019b; Hayes et al., 2020; Vakkuri & Abrahamsson, 2018). Hayes et al. (2020) also mentioned the possible impact of AI on ownership and property as a risk to take into account. These two risks were mentioned in philosophical literature, but usually were not subjected to in-depth philosophical discussions. A possible explanation for this is that these risks relate more to technical and legal aspects of AI than to ethical and philosophical issues.

2.3 Communication and AI

Before the 1990's many people believed that scientific findings and inventions would automatically lead to economic and societal advancements and members of the public were seen as passive innovation recipients (Schütz, Heidingsfelder, & Schraudner, 2019, p.129). This view has slowly shifted towards the aim that societal stakeholders should be involved in research, development and innovation (Schütz et al., 2019, p.129). A popular representation of the interaction between academic research and other societal actors is the quadruple helix, which was developed by Carayannis and Campbell (2009). The quadruple helix model shows how the four helices of academia/universities, industry, state/government and media-based and culture-based public intertwine to generate a national innovation system (Carayannis & Campbell, 2009, p. 206). Fraunhofer (2015) made an adaptation of the original quadruple helix model that looks at the helices from above, which can be seen in *figure 1*.

Figure 1

Quadruple helix model



Note. Quadruple helix model adapted by Fraunhofer (2015), originally developed by Carayannis and Campbell (2009).

The model by Fraunhofer (2015) in *figure 1* emphasizes that academia, industry, government and society are involved in multi-layered, dynamic, bi-directional interactions

(Schütz et al., 2019). Carayannis and Campbell (2009) argued that all four helices are equally important for the development of knowledge and innovation in the quadruple helix model. Though it is crucial to acknowledge that each of these four groups are involved, this research will focus mainly on the interaction between academic research and society. Therefore, these groups will be described in more detail below.

A report by the European MASIS expert group on the futures of science in society offers a more specific description of the stakeholders and social actors in research (Siune et al., 2009). Their categorization of stakeholders has considerable overlap with the groups in the quadruple helix model. However, they distinguished between researchers and academies on the one hand and schools and universities on the other hand (Siune et al., 2009, p. 21). In addition, they described media as a separate stakeholder group that has less interaction with researchers, but plays an important role in agenda-setting and the dissemination of research results into society (Siune et al., 2009, p. 24). Furthermore, they mention citizens as passive stakeholders, in the sense that clients are passive stakeholders of companies, since scientific developments have an effect on everybody in society even though they are not actively involved (Siune et al., 2009, p. 20). Later, they explain that citizens are usually only actively involved in science through their membership of other stakeholder groups (Siune et al., 2009, p. 23).

There are many reasons to engage citizens in science and technological developments. Fiorino (1990, p.226) argued that everyone in democratic societies has to cope with the effects of technologies and anticipate possible effects of new technologies. He argued that the risk assessment of new scientific and technological developments should not only be done from the perspective of risk professionals, but should include citizens to be more democratic (Fiorino, 1990, p. 227). Fiorino (1990) provided three arguments for this view. His first argument is the substantive argument that non-experts may find problems and solutions that experts miss and that their judgements are as sound as those of experts (Fiorino, 1990, p. 227). His second, normative argument is that according to democratic ideals citizens are the best judge of their own interests (Fiorino, 1990, p. 227). This argument is also present in the research agenda by the National Academy of Sciences (2017) which stated that it is important for people to receive information about developments in science and technology, since it can help them to make better decisions in different areas of their lives. The final, instrumental argument that Fiorino (1990, p.228) provided is that the participation of citizens leads to better results and makes risk decisions more legitimate.

Nevertheless, current developments in science and technology are complex and often relatively detached from society (Schäfer, 2017, p. 51). Because of this, most citizens receive information about science and technology mainly through news media (Schäfer, 2017, p. 51). Artificial intelligence (AI) is a good example of a technology in development that is complex, and which people need to know about, amongst others because it is expected that it will affect everyone in society. Walsh (2018) described that The World Economic Forum and multiple scientists argued that AI might drastically change people's

lives in various ways. Walsch (2018) focused mainly on economic risks of AI, for example the fear of AI taking over jobs, leading to high levels of unemployment. AI has been surrounded by speculations and fantasies about what it could be and become since early stages of its development (Natale & Ballatore, 2017, p. 4). These speculations and fantasies about AI centred around the belief that digital computing technologies can be seen as thinking machines (Natale & Ballatore, 2017, p. 4). The "AI myth" as Natale and Ballatore (2017) call this belief, had a large influence on the development of AI between the 1950s and the 1970s. In addition, the influence of this AI myth is still visible in the current narrative surrounding AI and related technologies (Natale & Ballatore, 2017, p. 13).

The study by Natale and Ballatore (2017) showed that communication about AI influences its development as well. This was also emphasized by Reinsborough (2017), who stated that interaction between scientific research agendas and public expectations has been important for the imagination of possible scientific futures. The interaction between scientific research and public understanding of that research does not always run smoothly. Especially in news reports about AI research there has been a lot of attention for what could go wrong in the development and application of AI, like AI becoming uncontrollable (Johnson & Verdicchio, 2017). This picture of AI is inaccurate according to experts in the field and can have a negative influence on the public understanding of AI (Johnson & Verdicchio, 2017). Hecht (2018) added that knowing what expectations and opinions citizens have about AI is important for developers of AI-technologies, even if they believe these citizens' views are unrealistic. According to Hecht (2018), AI developers need to be able to answer questions and address fears from citizens for their technologies to be successful. These studies also supported the instrumental argument Fiorino (1990) provided for involving citizens in the development and risk assessment of science and technology.

2.3.1 media reports about AI

There has been some scientific attention to how media report about artificial intelligence. Brennen, Howard and Nielsen (2018) conducted a media analysis of 760 reports about AI from six mainstream news outlets in the United Kingdom. They discovered that most news articles about AI discuss products, initiatives and announcements (Brennen et al., 2018, p.1). AI is usually portrayed as a solution to public problems (Brennen et al., 2018, p.1). In addition, Brennen et al. (2018, p.1) found that there is a difference between how right-leaning and left-leaning news outlets report about AI. Right-leaning outlets tend to focus on the influence AI might have on economics and geopolitics, whereas left-leaning news outlets pay more attention to ethical issues concerning AI (Brennen et al., 2018, p.1).

Chuan, Tsai and Cho (2019) conducted a similar study, focussing on the frames used in reports about AI in five major newspapers in the United States of America. For each of the newspaper reports they identified the prevalent topic, the type of impact framing (societal or personal) that was used, the type of issue framing (thematic or episodic) that was used and whether the report focused on risks or benefits of AI (Chuan et al., 2019,

p.341). They found that AI was predominantly discussed in relation to the topics of Business and Economy and Science and Technology (Chuan et al., 2019, p.341). Regarding the frames used, most articles used societal impact framing and episodic issue framing. There were slightly more articles that discussed benefits of AI than ones that discussed risks of AI (Chuan et al., 2019, p.342).

Chuan et al (2019, p.342) also analysed what types of risks and benefits were mentioned most often in the newspaper articles. The benefits they included were economic benefits, improving human life or well-being and the reduction of human biases or inequality (Chuan et al., 2019, p. 342). The types of risks they included were loss of jobs, shortcomings of the technology, unforeseen risks, runaway train, privacy, misuse, ethics and threat to human existence (Chuan et al., 2019, p. 342). The risks that were most frequently discussed in newspaper articles were shortcomings of the technology, loss of jobs and privacy concerns (Chuan et al., 2019, p. 342).

2.3.2 Framing

Some of the aforementioned studies looked into the framing of AI in newspaper articles. Framing is an important concept for this research as well. De Boer and Brennecke (2014, p.201) described framing as a multidimensional concept related to the production, content and effects of media messages. It has been used in various types of studies in different research areas within the social sciences and humanities (Entman, 1993, p.51). Entman (1993, p.52) provided an overarching definition of framing that is widely used: "To frame is to *select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation* for the item described." In short, frames emphasize certain aspects related to the topic that is being discussed whilst diminishing other aspects.

Framing presupposes that the way in which topics are presented in media outlets influences the way the public interprets these topics (De Boer & Brennecke, 2014). The definition by Entman (1993, p.52) implies that journalists deliberately use frames in order to persuade people of a certain view. However, this is not always the case. Journalists and editors may use frames in order to prioritize the information they include in news reports, De Boer and Brennecke (2014, p.206) called this process framebuilding. Once certain frames have been used in news reports this can influence the perspective on the topic that the audience adopts, which is called framesetting (De Boer & Brennecke, 2014, p.206). Through framebuilding and framesetting frames can be used and adopted deliberately, but they can also be used unintendedly or lead to other effects than frame adoption (De Boer & Brennecke, 2014, p.206).

Several frames have already been identified in the news coverage of AI in previous studies. Chuan et al. (2019) distinguished between frames and topics that were present in newspaper articles about AI. They focused on the broad frames of risk and benefit framing, personal and societal impact framing, and episodic and thematic framing. Brennen et al.

(2018) mentioned that they identified frames, topics and recurring themes in newspaper articles about AI, but seem to use these terms interchangeably. Because of this it is not clear what frames they identified.

2.3.3 The role of experts in science communication

The media analysis by Brennen et al. (2018, p. 4) that was discussed in the previous sections showed that most newspaper articles about AI were framed around industry products. In addition, Brennen et al. (2018, p. 1) looked at which experts were mentioned in the articles. They found that one-third of the unique sources mentioned in news articles about AI were people affiliated with industry (Brennen et al., 2018, p. 4). Approximately another third of the unique sources mentioned consisted of quotes from written sources such as press releases and official statements (Brennen et al., 2018, p. 4). Among the rest of the unique sources mentioned, approximately 17 percent were connected to academic institutions, 5 percent to governmental and political organizations and 3 percent to advocacy organizations (Brennen et al., 2018, p. 4). Based on these findings Brennen et al. (2018, p. 9) recommended that newspapers should include a more diverse range of sources in articles about AI, including experts from different fields and citizens. Chuan et al. (2019, p. 342) obtained similar results in their media analysis, which showed that 64,7% of the sources mentioned were people associated with industry, followed by 29,1% consisting of scientists and 23,6% other experts.

Following up on their earlier research, Brennen, Schulz, Howard and Nielsen (2019) examined more closely which academic experts were mentioned most often in newspaper articles about AI. They identified the 150 most-cited academic scholars in the field of AI in Google Scholar and looked at how often they were mentioned in articles in major newspapers in the UK and USA (Brennen et al., 2019, p. 2). They found that the 10 researchers that were mentioned most often, made up 70% of all news mentions in the sample (Brennen et al., 2019, p. 4). Additionally, the researchers with the most citations in Google Scholar were usually not the ones who were mentioned most often in newspaper articles (Brennen et al., 2019, p. 4). Instead, Brennen et al. (2019, p. 4) found that researchers who had industry affiliations as well as academic affiliations were mentioned most often in newspaper articles. Industry-affiliated researchers accounted for 56,6,% of news mentions and 15% of Google Scholar citations in the UK and for 71,9% of news mentions and 19,3% of Google Scholar Citations in the USA (Brennen et al., 2019, p. 4). This shows that even when newspaper articles mention academic researchers, these researchers are often connected to industry as well.

2.3.4 The role of the public in science communications

One study has been published that focused on the perceptions of Dutch citizens about AI and communication about AI. The Dutch Ministry of the Interior and Kingdom Relations commissioned Kantar Public to research what perceptions Dutch citizens have about AI and possible governmental use of AI (Verhue & Mol, 2018). They first organized two group discussions with 16 people in total to get an understanding of citizen's first associations

with AI and what risks and benefits they anticipate AI to have (Schothorst & Verhue, 2018, p. 1). When the participants were asked what they thought about when they heard the term “artificial intelligence”, computers, robots, science fiction and some possible applications of AI, like speech recognition and autonomous cars were mentioned in both groups (Schothorst & Verhue, 2018, p. 3). In the group with low skilled participants most associations were related to hardware and automation (Schothorst & Verhue, 2018, p. 4). In the group with highly educated participants most participants had some understanding of what AI was and some already mentioned possible societal implications (Schothorst & Verhue, 2018, p. 4). However, most participants in both groups had trouble explaining what AI is (Schothorst & Verhue, 2018, p. 4).

After the researchers explained what AI and machine learning is, both groups asked questions about the boundaries of AI and automation and expressed fears of a lack of human control over AI systems (Schothorst & Verhue, 2018, pp. 4–6). Nevertheless, most participants in both groups did not worry about AI a lot, since it is not visible in their daily lives (Schothorst & Verhue, 2018, p. 5). When asked about possible negative applications of AI the highly educated participants mentioned that using AI in jurisdiction could lead to a lack of human measurements, emotion and control (Schothorst & Verhue, 2018, p. 8). The participants in the other group had trouble imagining specific possible applications, but feared that it could cause people to lose their job and that it could reduce their privacy (Schothorst & Verhue, 2018, p. 8). Possible applications of AI the participants would approve of mainly included applications in the areas of medicine, crime prevention, dieting, marketing and route planning (Schothorst & Verhue, 2018, p. 7).

3. Methods

In order to answer the research questions, three separate studies were conducted. Firstly, semi-structured interviews were conducted with experts in the field of artificial intelligence (AI), in order to get an overview of the current state of development of AI and of the expectations that experts have about the future of AI and its applications. Secondly, a media analysis was conducted to discover how news media report about AI. Finally, focus group interviews were conducted with members of the public, to find out what expectations they have about AI and what these expectations are based on. Since it can be hard for people to understand what AI is and what applications it might have, newspaper articles were used as scenarios to make it easier for the participants to discuss AI.

3.1 Expert interviews

In order to investigate what expectations experts in AI have about developments in this field, semi-structured interviews with experts in AI were conducted. The aim of these interviews was to get an overview of the current state of development of AI, as well as of the experts opinions on the impact of AI on society and their expectations for the near future of AI. The interviews took approximately 45 minutes and were conducted via online videoconferencing tools, such as Zoom and Google Meet. Before the interviews, the participants were asked for their consent. The interviews were recorded, transcribed and pseudonymized for further analysis. Ethical approval for the study was obtained in advance.

An interview protocol with the main questions was used as a base for the semi-structured interviews. The interview protocol can be found in appendix A. Depending on the answers of the participants, follow-up questions were asked in order to get more complete and in-depth answers. Each interview started with a short introduction about the research and the procedure of the interview. After this introduction, the participants were asked about what their work is and how it involves AI. Following this, they were asked to tell something about the current state of development of AI and what their expectations are for the future of AI. After this more general part of the interview, they were asked to discuss possible societal impacts they think their work and the AI they work with might have. This included questions about whether their work incorporates any customs or procedures that draw attention to ethical and social implications of their work. The final part focused on communication about AI towards the public. Participants were asked if they are involved in communicating with laypeople about AI and what they think about the way AI is portrayed in news media.

A convenience sampling strategy was used to recruit the participants. The participants were recruited via the personal network of the researcher and via searching through members of interest groups related to AI. The inclusion criteria for participants to take part in the interviews were that they had to work with AI in the Netherlands and they had to be able to speak Dutch. The interviewees were selected to represent the three expert groups of governance, academic research and industry from the quadruple helix as

discussed by Carayannis and Campbell (2009). In total, six participants were included, two for each of these categories. An overview of the participants, their area of expertise, educational background and gender can be found in table 1.

Table 1

Overview of participants expert interviews

Participant	Category	Area of expertise	Educational background	Gender
1	Industry	Computer vision	Applied physics	Male
2	Governance	Public debate	Philosophy	Female
3	Academia	Search engine algorithms	Computer science	Male
4	Academia	Medical AI	Mathematics and physics	Male
5	Governance	Organizational change	Political science	Female
6	Industry	Data science	Social science and data science	Female

To analyse the interviews, a codebook was created in an iterative process with the researcher and the supervisors. The codebook was based on the concepts discussed in the literature review and the questions in the interview scheme. Open coding was used to include more specific ethical and societal implications and other recurring topics. Since the experts were asked to provide an explanation of their work and of the technologies they worked with, codes for explanations, examples and sources that they mentioned were included as well. Finally, codes about the attitudes participants had towards AI were included in the codebook. For positive attitudes the codes of benefit, hope, affordance and promise were used. The codes of risk, fear and limitation were used to analyse negative attitudes. The codebook can be found in appendix B.

3.2 Media Analysis

A media analysis was conducted to address the research questions “How do news media report about AI?” and “What considerations about the societal impact of AI are apparent in the public debate about AI?”, focussing on newspaper articles. The database NexisUni was used to search for Dutch newspaper reports about AI. The search was limited to newspaper articles that were published between September 1st 2019 and August 31st 2020. Similar newspaper articles were grouped together using a filter from NexisUni, so the same article would not show up twice in the results.

Searching for the term “kunstmatige intelligentie”, the Dutch translation for artificial intelligence, resulted in 2102 individual newspaper articles, 828 of these articles were published in the main national newspapers: Volkskrant, NRC handelsblad and NRC next, Telegraaf, Het Financieele Dagblad, Trouw and Nederlands Dagblad. Searching for the term “machine learning”, resulted in 86 individual newspaper articles, 55 of which were

published in the main national newspapers, 45 of these articles also included the term “kunstmatige intelligentie”. The 828 articles from the main Dutch newspapers which included “kunstmatige intelligentie” were selected for a large scale analysis. This sample was chosen since the majority of the articles found by the term “machine learning” was also included in this sample.

After trying various methods to create a representative sample, in total, 53 of the 828 newspaper articles were selected for an in-depth content analysis. These articles were sorted based on relevance through the algorithm of NexisUni, which is partly based on how often and where in the article the search term is mentioned. The articles from the first 8 pages with the most relevant results were downloaded for further selection. All articles with 500 words or less were removed from this sample. A few articles were removed because they only mentioned AI as a small example in a discussion about a different topic or in the context of fiction that was reviewed. This lead to the total sample of 53 newspaper articles, the references for these articles can be found in appendix C.

In order to analyse recurring themes in the newspaper articles, a codebook was created. The first version of the codebook was based on the theoretical framework and included code groups for the newspapers and sections the articles were published in, the sources mentioned in the articles, the frames that were used and the philosophical concepts that were mentioned. The codes for the newspaper, section and type of articles were coded on article level. The sources and philosophical concepts that were mentioned were coded on a sentence level. Of the frames that were put forward by Chuan et al. (2019), the impact and issue frames were coded on the article level and the risk and benefit frames were coded on the sentence level, since some newspaper articles discussed both risks and benefits of AI. This codebook was adapted through an iterative process of coding the first few articles and adding open codes for new themes within the aforementioned categories and other recurring topics. The final version of the codebook can be found in appendix D.

The reliability of this codebook was assessed by calculating the intercoder reliability. From the sample of 53 newspaper articles, ten articles were randomly selected to be coded by a second coder. This selection process falls within the 10-25% margin of data units as recommended by O’Connor and Joffe (2020, p. 6). First, the second coder received the codebook and an explanation of the categories and codes and how to apply them on article or sentence level. Secondly, both coders independently coded one of the ten articles and discussed the process to clear up any confusions about specific codes afterwards. Following this, both coders independently reassessed the coding of the first article and coded the rest of the ten articles. Finally, the intercoder reliability was calculated using the Krippendorff’s Alpha measurement as implemented in Atlas.ti. The main advantage of this measure is that it allows for multiple codes to be applied to the same or overlapping pieces of text (Friese, 2020). The cumulative Krippendorff’s alpha for all codes and all ten articles was 0,811, which is above the recommended minimum of 0,8.

3.3 Focus group interviews

The final study consisted of focus groups to investigate what knowledge and expectations Dutch citizens, with no professional experience with AI, have about AI and its possible societal impact. The focus groups were conducted via the online videoconferencing tool Google Meet and took approximately one hour. Each focus group session was recorded, transcribed and pseudonymized for further analysis. Before the focus groups, the participants were asked for their consent and ethical approval for the study was obtained in advance.

An interview protocol with a list of topics and questions was used to structure the focus groups and keep track of time, this protocol can be found in Appendix E. The topics and questions were based on the recurring topics from the expert interviews, media analysis and literature review. The report of the focus group study about AI by Schothorst & Verhue (2018) was used as an example for the structure of the focus group study. After an introduction round, the participants were asked about their current knowledge of AI and their use of news media and social media. Following this, the researcher provided an explanation of AI and some examples of applications. The participants were asked to what extent this explanation matched what they thought about AI before and what benefits and risks of AI they could think of.

After the general questions, two fragments from newspaper articles about AI were discussed in each focus group. The newspaper articles provided examples of specific applications of AI as well as an explanation of how they are used in context. This allowed the participants to have an in-depth discussion based on a shared understanding of a specific application of AI. Four newspaper articles were selected from the sample of the in-depth media analysis. Two articles with a more positive attitude towards AI and two articles with a more cautious or negative attitude towards AI were selected. These articles were summarized in order to highlight the relevant discussion points and reduce the time participants needed to spend on reading the articles.

The summaries of the news articles were included in the focus group protocol. During focus groups 1 and 3 the first two news fragments were used and during focus groups 2 and 4 the last two news fragments were used, so that both positive and negative aspects of AI were highlighted in each focus group. For each news fragment the participants were asked about their initial reaction, the risks and benefits and the possible societal impact of the AI application that was described. The participants were also asked about how reliable they thought the newspaper article was. After the discussion of the news fragments the participants were asked if their views of AI had changed and what impact they thought AI has or might have on their own life. Finally, the participants were asked about their expectations for further developments in AI in the coming five years.

Ideally, there would have been three focus groups with six participants in each group (Guest et al., 2017). However, due to the measures to prevent the spread of the Covid-19 pandemic the focus groups had to be held online. To facilitate a smooth discussion

in an online setting the amount of participants per session was reduced to a maximum of five and an additional focus group was organized. Two focus groups with four participants and two with five participants were organized. The participants were recruited from the personal network of the researcher through snowballing and the groups were created based on the availability of the participants. An overview of the demographic characteristics of the participants per focus group can be found in table 2.

Table 2

Overview of the participants per focus group

Focus group	1	2	3	4	Total
Number of participants	4	5	5	4	18
Male	2	3	2	1	8
Female	2	2	3	3	10
Higher educated (HBO-WO)	3	5	5	2	15
Lower educated (MBO)	1	0	0	2	3
Age range	19-55	24-51	23-29	25-59	19-59

For the analysis of the focus groups, a codebook was created in an iterative process with the researcher and the supervisors. The codebook for the expert interviews was used as a basis and adapted to fit the protocol for the focus groups. The codebook for the focus groups included the concepts from the literature review as well as more specific concepts and recurring themes that emerged in the expert interviews and media analysis. Since the participants were asked about their knowledge of AI and its impact, the codes about explanations, examples and evaluations of AI were included as well. In addition to the codes about sources that were mentioned, the concept of trust was added, since the participants were asked how much they trusted various stakeholders. The codebook for the focus groups can be found in appendix F.

4. Results

This chapter will discuss the results of the three empirical studies that were conducted. Firstly, the results of the interviews with experts in the field of artificial intelligence (AI) will be discussed, focusing on the experts explanations of AI, their role in and perspective on communication about AI and their expectations for the impact of AI on society. Secondly, the results of the large-scale and in-depth media analyses of newspaper articles about AI will be discussed. Thirdly, the results of the focus groups will be discussed, focussing on the participants' pre-existing knowledge of AI, their opinions on communication about AI and their expectations for the impact of AI. Finally, the results of the three studies will be compared to each other.

4.1 Results from the expert interviews

4.1.1 Explanation of AI

Since there is no widespread consensus about the exact definition of AI and machine learning, all six participants in the expert interviews were asked to describe these concepts. Five of the participants found it difficult to give an immediate, clear definition, stating that it was difficult or "a good question". Participants 2, 5 and 6 also mentioned they had noticed there is disagreement among people about what they mean when using the terms AI and machine learning. Participant 5 explained that she always uses the definition of artificial intelligence from the Dutch government, but that she could not learn it by heart. The definition from the Dutch government is a translation of the definition given by the High-Level Expert Group on Artificial Intelligence (AI HLEG, 2019): "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals".

Participant 6 provided the shortest definition of machine learning, stating that for her machine learning is "an algorithm, so a piece of code, that can continue to learn by itself". The self-learning aspect of machine learning algorithms was an important part of the definition for most participants. Participants 2 and 4 both mentioned that pattern recognition is an important part of machine learning, since the computer learns to detect patterns in data by itself. A few participants compared machine learning algorithms to other types of algorithms by focussing on the role of programmers. For example, participant 3 explained: "If you don't use machine learning as a programmer, you have to decide how the system makes decisions (…) and with machine learning we try to leave as much as possible to the system itself by giving it examples".

When asked to compare the different concepts, most participants saw AI as a broad field of technologies, of which machine learning is a more specific part. For example, participant 4 stated: "I would put them hierarchical. Artificial intelligence is like an umbrella, and machine learning is a part of that". Some participants also mentioned other technologies and categories related to AI, like general AI, robotics, data science and different types of machine learning, including deep learning and strategies like supervised,

unsupervised and reinforcement learning. Some participants mentioned examples of applications of AI and machine learning to clarify their explanation. For example, participant 2 used the example of Alpha-Go and Alpha-Zero to explain the difference between machine learning and other algorithms: “Alpha-Go has been trained based on the games played by people, whereas Alpha-Zero learns by itself, by playing against itself”. Participant 5 mentioned that she uses different examples to explain AI to different audiences, like comparing algorithms to cooking recipes, describing chess computers or more difficult analogies and formulas.

The participants were also asked what current possibilities of AI and machine learning they found most impressive. Many of them mentioned specific applications, like the use of AI in healthcare, the option for AI applications to take over dull, dirty and dangerous work, the ability of a machine learning algorithm to beat the game “Go” and to outperform traditional models for search engines. Participant 1 mentioned a translation algorithm that could be used do many different things: “Most neural networks are relatively specialized, but this one could translate Dutch to English for example (…) but it could also create a website based on a written description”. This algorithm seemed to take a small step towards general AI, which aims to create artificial intelligence that can outperform humans in many different areas. Participant 1 found this impressive, but also a bit objectionable.

4.1.2 Communication about AI

During the interviews, the participants answered a few questions related to communication about AI. For five of the six participants talking to others about AI is a part of their job. For participants 5 and 6 this mainly consisted of helping colleagues within their organization or partner organizations to understand and implement AI applications in a responsible way. Participants 3 and 4 both teach and supervise students as part of their work and participant 6 also teaches students in addition to her main job. She explained: “I teach classes and workshops at another organization and I mentor a few people (…) technically and in soft skills”.

Both participants 2 and 4 said they play a role in directly or indirectly informing laypeople about AI. Participant 4 explained that the organization he works for recently published a video of him explaining what they do with AI and how it works. Participant 2 explained that she usually writes a long report first and draws from that for shorter articles and presentations: “Once the report has been published, we also write more accessible pieces for the website or for specific magazines or articles. We also try to share our story in presentations, or in debates or during events”. Participants 3 and 6 both said that they took various opportunities to inform laypeople about AI during events or online during their free time. Participant 3 stated: “I try to take these opportunities every now and again when they arise”. For participant 1 informing others about AI was not part of his job and he did not speak about it in public either, but he sometimes discussed new developments in AI with friends and colleagues.

When asked about the sources they use to get information about AI, most of the participants (5 out of 6) mentioned scientific journals and trade publications as an important source of information. Many participants also received information through speaking with colleagues and attending conferences and other events. In addition, some participants found information about AI in news media and on social media. For example, participant 6 explained that she followed specific people to stay up to date: “Twitter is the best way to follow people who just retweet all of the new things that happen”. Participants 2 and 5 both referred to the Dutch National AI Course as a good place for people to start learning about AI. A few specific popular books and tv shows about AI applications were mentioned as well, like the book “Weapons of Math Destruction” by Cathy O’Neil, and Netflix’s docudrama “The Social Dilemma”. Participants 2 and 4 both mentioned the Dutch talk show “Zondag met Lubach”, talking about a segment on apps to prevent the spread of Covid-19 and a segment on how conspiracy theories spread via social media.

Four of the six participants explicitly stated that there is almost too much information about AI, which can make it hard for people to find what they want to know. Participant 2 explained about this as follows: “It’s not that there is a lack of information, the problem is often to make it accessible, easy to find and understand”. This overflow of information was also described as a hype surrounding AI. Participant 3 said about this: “At the moment there is a bit of a hype in the area of AI and machine learning.” He also stated that he expected this hype would be over relatively quickly.

The participants had mixed opinions about how news media represent AI. Most participants were happy that news media pay attention to the topic and discuss both positive and negative aspects of AI. Participant 1 did not follow the news, but thought AI was mainly discussed in a critical way, explaining: “They especially have big questions about how far it can all go. I also find that hard to estimate, so for people outside of the field that is even more difficult of course”. Participant 5 thought there was a good balance between positive and negative articles, but commented that “it depends on what newspaper you read”. Participant 6 thought AI was discussed in a positive light more often than in a negative light, explaining: “I think the explicit articles that say “we’re going to talk about AI” are almost always critical (…), but indirect articles, where AI is a small component, are always positive”.

More participants shared this sentiment that news articles often lack nuance in their representation of AI. Participant 3 provided an example of this: “What you often see is that they say it’s very bad that Google can predict exactly what we want to buy with those advertisements. But then I think, well but Google can’t do that at all”. Participant 4 thought this lack of nuance was somewhat unavoidable, stating: “They try to do justice to the research, but they have to write in a way that’s suitable for a broader audience, because of which many of those nuances are lost”.

4.1.3 Impact of AI

When asked about the impact of AI on society, all participants of the expert interviews mentioned both advantages and disadvantages of AI. Specific applications of AI and their impact were usually discussed in a nuanced way, with participants mentioning both positive and negative aspects. The positive attitudes of benefit, hope, affordance and promise occurred 26 times in total. All of the participants expressed some positive expectations about the development of AI in the near future. For example, participant 6 stated: "The theory [of machine learning] will make enormous steps, because of the popularity and the funding it currently has, but also because the computing power increases so quickly".

The negative attitudes towards AI, including limitations, risks and fears, occurred in 30 quotes in total, 18 of which were about risks of AI. An example of a risk that was mentioned was the scalability of machine learning models, as participant 3 explained: "Once you have such a system you can apply it to millions of people. I think that's very dangerous, because the people for whom a mistake is made don't have any options to correct that mistake". In addition to these general attitudes, the participants discussed particular examples of benefits and risks related to AI. This included explicit mentions of the philosophical concepts discussed in the theoretical framework. Table 3 shows how often each of these concepts were discussed and provides examples of quotes for each concept.

Table 3

Philosophical concepts in expert interviews

Category	Frequency	Quotation
Autonomy	2 times by 1 participant	P2: (talking about AI applications for HR): "It affects privacy, autonomy, it doesn't eradicate discrimination at all."
Bias	8 times by 5 participants	P4: "You have that classic example that they try to predict the risk of recidivism and that model has a bias from the data it was built with, so it attaches too much value to ethnicity instead of to the actual risk of recidivism."
Explainability	7 times by 5 participants	P3: "Those systems are black boxes, we don't know very well why they work."
Fairness	1 time by 1 participant	P2: "These ethical codes often link to the traditional bio-ethical principles, (…) including justice, fairness, beneficence and non-maleficence."
Responsibility	6 times by 3 participants	P5: "I feel responsible to (…) help people to make an informed decision about whether or not to use AI, and if we do it than we do it extremely decently"

Privacy	5 times by 3 participants	P1 (talking about ethical procedures in work): (…)The area of privacy is a very clear focus area, but ethics is not really.

As table 3 shows, all of the philosophical concepts from the theoretical framework were mentioned at least once. However, in many cases the participants did not directly refer to these concepts, but talked about more noticeable topics. For example, even though only one participant mentioned autonomy directly, five participants talked about the relationship between people and AI. Participant 5 explained how experts in her organization use AI, stating: "We use it as additional information, not to press buttons, but just as additional information that professional people use to do their job". This quote shows that the participant does not see the AI application as a fully autonomous actor, the person using the application remains in control of what happens. This also indicates that the responsibility is attributed to the person using the application.

The theme of responsibility was touched upon multiple times when participants discussed the relationship between people and AI, the reliability of AI applications and the education of people making AI applications. For example, participant 3 stated: "I hope that our students at least realize that they have a large responsibility once they graduate". Participant 6 also argued that data scientists should be responsible for the algorithms they create and the biases it could include and added that this was missing in her education: "It isn't taught to you when you become a data scientist that you should take this into account. There are toolkits to measure or correct for inequality whilst you are making it".

Even though the concept of fairness was only mentioned once, the topics of power differences, discrimination, prejudice and diversity recurred often. Participant 2 talked about prejudice in algorithms that employers can use in job application processes, explaining: "The companies that use these technologies think I don't want to discriminate, I want to remove that human bias […], but they often don't realize how discrimination can sneak into your dataset or into the way your algorithm trains in many different ways". As this quote demonstrates, these discussions were often related to bias in data and algorithms as well, since such biases regularly result in prejudice and discrimination. Participants 5 and 6 both proposed that including diverse perspectives in the process of making AI applications could help to prevent the inclusion of undesirable biases and prejudice in algorithms.

The concept of explainability was discussed in relation to the transparency of algorithms a few times. For example, participant 1 stated: "It is hard to explain, since not everything has completely been decided by a programmer, because a machine learning algorithm learns by itselfs. After a while, with a few outcomes you wonder if that is caused because the algorithms learned it that way or if the company or programmer meant to do it that way". There was some disagreement about this topic among the participants. Most participants saw the lack of transparency and explainability of machine learning algorithms as a potential problem, but participant 4 thought there was a bit too much negative

attention for this issue. He argued that news media treat AI as " a sort of magical thing, as if it's a black box that can do anything. But that's not true at all, it is a relatively simple type of models".

One other concept that was discussed in the theoretical framework and recurred in the interviews was the impact of AI on climate change. Interestingly, during the interviews AI was discussed as contibuting both to the cause and to possible solutions of climate change. Participant 4 discussed the negative impact AI can have on climate change as follows: "You sometimes see those pictures that show it's ten circles around the earth with a Boeing 747 to train such a model, in CO2 emmissions". On the other hand, participants 2 and 5 both mentioned applications of AI that could help to reduce climate change. Participant 2 explained: "For example in the energy transition they are working on datagovernance and how things like smart meters and electric cars can be used to make the system more flexible and achieve the goals in the climate agreement".

When discussing the impact of AI on society, five of the six participants talked about politics, law and regulation of AI. Many participants noted that there is an increase in the attention for the impact of AI in politics and regulation and were positive about this trend. Participant 2 argued that governments might have given companies too much room to innovate, stating: "They often say that politics and regulation always lag behind technological developments, but that's a choice. It is important to not only focus on the technology, but also work on social, economic and legal innovations that are needed to embed it". In contrary, participant 4 thought European governments risk underinvesting in AI: "From my viewpoint they partly hold back innovation based on arguments that are not always correct or overestimate the possibilities". Four of the participants also expressed worries about the increasing power of large technology companies. For example, participant 3 stated: "The companies like Google, Facebook and Amazon have so much power, and you see that they get more and more power through laws and regulations".

Another recurring topic was how algorithms influence the spread of information, with participants mentioning targeted marketing, filter bubbles and the spead of disinformation and fake news. Many participants mentioned targeted advertising and algorithmic reccomendations on social media platforms as examples of the current impact of AI on society. For example, participant 1 said: "If you look at something like YouTube for example, that the suggestions you get there are decided by an algorithm, I think most people are aware of that". Filter bubbles and the spread of disinformation were often discussed together as a risk of AI, as participant 5 stated: "Ofcourse I think I'm in a very sensible bubble, but people in a conspiracy theory bubble probably think that as well".

4.2 Results from the Media analysis

4.2.1 Large scale analysis

Searching for the term “kunstmatige intelligentie”, the Dutch translation for artificial intelligence, resulted in 2102 individual newspaper articles, of which 825 were published in the main national newspapers. Duplicates of articles were filtered out by NexisUni, which was based on a percentage of similarity between articles and led to 825 unique articles. The division of articles among the main Dutch newspapers is shown in table 4.

Table 4

Division of articles per newspaper

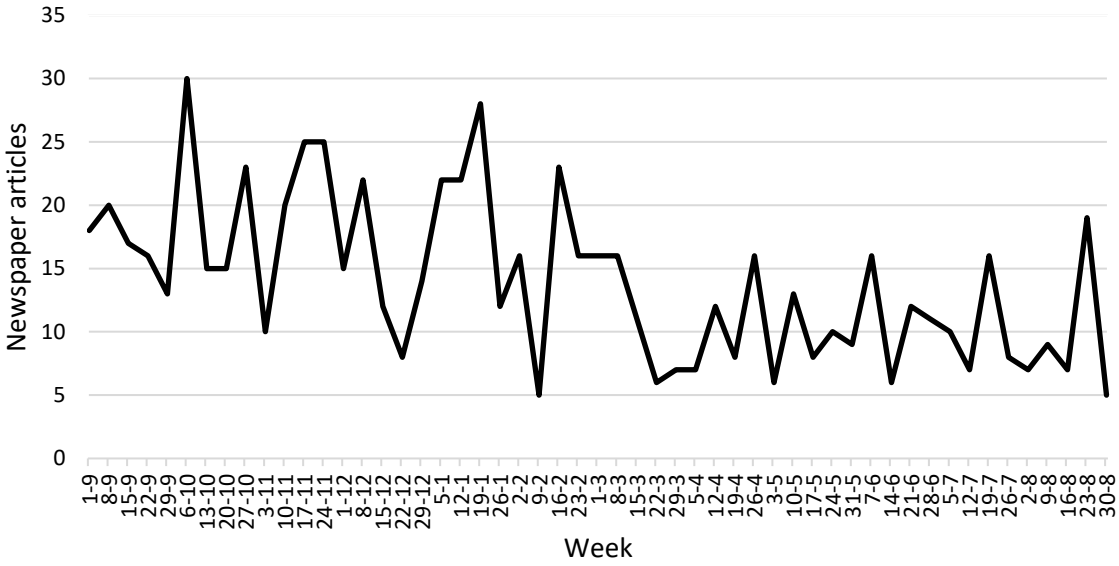
Newspaper	Number of unique articles
Het Financieele Dagblad	249
NRC Next	135
NRC Handelsblad	131
De Volkskrant	108
Trouw	78
De Telegraaf	64
Nederlands Dagblad	60

The timeline in *figure 2* shows how many articles were published in each week from September 1st 2019 until August 31st 2020. The horizontal axis shows the starting date of each week in the sample and the vertical axis shows the number of newspaper articles. It should be noted that this timeline only includes 740 out of the 825 newspaper articles. This seems to be caused by the sensitivity of the filter that groups duplicates of articles in Nexisuni. For the analysis of the timeline, the articles were selected per week instead of per year, which caused more articles to be grouped together as similar articles.

Figure 2 shows that the first large peak in the publication of news articles about AI occurred in the week of October 6th, in which a lot of news articles discussed the National AI strategy that the Dutch government announced. In the week of December 22nd only 8 articles about AI were published, which is probably because it was Christmas during that week. In the week of December 29th there were a lot of articles that looked back on the past decade or forward to the new decade, which mentioned developments in AI as part of this. The second biggest peak occurred in the week of January 19th, when many articles discussed news about data management issues in banking companies.

Figure 2

Timeline of publication of newspaper articles per week



In the week of February 9th only 5 articles mentioned AI, this is probably due to newspapers focussing on the coronavirus pandemic around that time. Nevertheless, there was a peak in the number of articles in the week of February 16th, most of these articles discussed the new plans the EU published about the development of AI. The smaller peaks in the weeks of June 7th and July 19th both included multiple news articles about racism and how algorithms might contribute to that. In the final peak in the week of August 23rd there were ten newspaper articles about a collaboration between the two universities in Amsterdam and Huawei that caused some controversy. The final week in this graph only consists of 2 days, in which 5 newspaper articles about AI were published. In the full week that started on August 30th 12 newspaper articles about AI were published.

4.2.2 In-depth analysis

For the in-depth analysis a smaller sample of 53 newspaper articles was selected from the newspaper articles that were included in the large scale analysis. Table 5 provides an overview of the articles divided per month and the newspaper they were published in. As table 5 shows, the division of articles per newspaper for the sample of the in-depth analysis is similar to that of the large-scale analysis, considering that there was considerable overlap between the newspaper articles in NRC Handelsblad and NRC Next. The peaks in the number of newspaper articles about AI in October, January, June, July and August in the timeline of the large scale analysis in Figure 2 are visible in the sample for the small scale analysis as well.

Table 5

Division of articles per month and per newspaper

	Financieele Dagblad	NRC	Volkskrant	Trouw	Nederlands Dagblad	Telegraaf	Total
September	1	1					2
October	2	3	1			1	7
November					1		1
December	2		1				3
January	1	2		1	2		6
February	1		1				2
March	2	1					3
April							0
May	2						2
June	5						5
July	3	2		2			7
August	7	2	2	2	1	1	15
Total	26	11	5	5	4	2	53

4.2.2.1 Frames, sources and recurring topics

The first step of the in-depth media analysis focused on how AI is represented in Dutch newspaper articles. For each newspaper article, issue frames and impact frames were identified. A majority of the newspaper articles (30) was framed around episodic issues, but there were 23 articles that focused on thematic issues too. The articles about episodic issues mainly discussed new applications of AI, breakthroughs in the development of AI and news about changes in funding and regulations of AI. The articles that discussed thematic issues on broader trends in the development of AI and the impact of AI on society.

Most newspaper articles (35) focused on the societal impact of AI. These articles usually discussed how an application of AI might impact citizens in general, for example through discrimination in decision-making algorithms. Sixteen articles were framed around the impact AI has on a specific group, like specific industries or groups within society, such as elderly people. Only two of the newspaper articles were framed around individual impact. One of these articles was an opinion article and the other one was a review of an exhibition. Besides this, there were more articles that mentioned the impact of AI on an individual as an example of the impact on a larger group or society.

The frames of risk and benefit and the affective reactions of hope and fear of AI were coded per sentence. Table 6 shows how often each of these codes occurred. Most articles seemed to have a negative attitude towards AI, with risks and fears occurring a lot more often than hopes and benefits. Interestingly, hope and fear were often mentioned together in the same articles. These two codes even co-occurred in the same sentence six times in two different articles, as exemplified by the quote for hope in table 6. Even though

hope and fear or risks and benefits were sometimes mentioned together in the same article, most newspaper articles had either a positive or a negative focus.

Table 6

Evaluation of AI in newspaper articles

Category	Frequency	Quotation
Benefit	44	Feb1: "Artificial intelligence offers great possibilities and we have to unleash it's potential."
Risk	79	Jul7: "Technology becomes increasingly complex, which causes the risks to become more complicated."
Hope	7	Dec2: "Artificial intelligence brings us just as much hope as fear."
Fear	17	Nov1: "There are many visions of fear surrounding AI: What if the computer autonomously develops into something we don't want, like robots that see humans as subordinate and destroy them?"

When talking about specific applications of AI, positive aspects were mentioned a lot more often than negative aspects. Affordances of AI, what it can do and what it has made possible, occurred 97 times in the newspaper articles. An affordance that was mentioned in several articles was the ability of AI to accurately analyse a large amount of pictures. For example, one article (Dec1) quoted a professor, who said: "Everywhere where you let people look at photographs, you can also use computers". In addition to this, promises of what AI might be able to do in the future occurred 58 times in the newspaper articles. Conversely, limitations of AI applications occurred 41 times in total. These limitations included areas of AI that need further development before they work properly and discussions of tasks that are very difficult for AI in comparison to how difficult they are for people, like being creative and having smooth conversations.

Most newspaper articles mentioned sources that provided information about AI. This included people that were interviewed for newspaper articles, as well as organizations and individuals that were mentioned because they played a role in the development or regulation of AI or because they were impacted by the use of AI. As table 7 shows, most sources that were mentioned were affiliated with industry, followed by academia, politics and governance and interest groups. Even though multiple articles talked about AI making art or works of art related to AI, artists were mentioned as sources the least often, followed by citizens and references to other media.

When these sources are compared to the quadruple helix, it is clear that all four helices appear in newspaper articles about AI. The categories of sources that were mentioned less often, which were interest groups, media, citizens and artists, all represent different aspects of society in the division of the quadruple helix. Together, these groups were mentioned 47 times, which is almost as often as sources affiliated with governance.

This shows that newspaper articles mainly focus on actors in the groups of business and academic research, who are usually most actively involved with the development of artificial intelligence, but there is also place for discussions about AI focussed on governance and society.

Table 7

Sources mentioned in newspaper articles

Category	Frequency	Quotation
Academia	61	May2: "The University of Utrecht helps medical personnel to find the proper corona treatment with AI."
Artists	4	Dec1: "Artist James Bridle mocked systems that can't do anything but obediently following rules."
Citizens	7	Oct1: "Alice is standing on the table of Johanna de Boer (92) in a nursing home in Akersloot."
Industry	77	Aug10: "Marc Hesselink, analist at ING thought the takeover was "surprising"."
Interest groups	27	Jul2: "The Consumers Association also warns for "un-insurability"."
Media	9	Jul3: "The Times came up with the virtual "news butler" James, that presented news based on earlier clicks."
Governance	48	Feb1: "These are the main concepts in the plans of the European Commission with artificial intelligence and data, that were published yesterday."

In addition to the frames and sources that were mentioned, there were some recurring themes and topics that did not directly relate to the impact of AI. These topics cover areas of application of AI, like healthcare and games as well as trends in society that have an effect on the development of AI, like climate change, the Covid-19 pandemic, geopolitics and regulations. An overview of the topics and how often they were mentioned can be found in table 8. The topics of geopolitics and regulation, which recurred most often, were regularly discussed in relation to each other as well. For example, the European privacy regulation was often mentioned as an example when the development of AI in Europe was compared to developments in other countries and continents. A quote that illustrates this is: "It is relevant for the whole world if Europe sets boundaries for products. Take the European privacy regulation GDPR, we already see that being taken over by American states" (Oct4).

The theme of fake news and disinformation was discussed three times in the newspaper articles and all of these times it was related to regulation. One article (Mar3) discussed whether governments should restrict social media like Facebook in their countries, because they spread misinformation. The other article in which these topics were discussed focused on the responsibility of companies like Facebook to prevent the spread of fake news and misinformation. This article (Jul5) stated: "A few weeks ago

Facebook-chef Mark Zuckerberg said proudly that he did not want to be a “referee of the truth”. This caricature distracts from the question if he wants to protect his customers from virtual hooligans and if he thinks anti-discrimination laws also apply to Facebook”.

Table 8

Recurring themes in newspaper articles

Category	Frequency	Quotation
Climate change	17	Aug9: “Thanks to machine learning we can reduce our energy use enormously by increasing the efficiency of how we use energy.”
Corona virus	5	Mar2: “Governments and companies employ artificial intelligence as a weapon against Covid-19.”
Games	7	Nov1: “After checkers and chess a computer can now beat people in a strategy game.”
Geopolitics	57	Jul7: “The USA is putting pressure on allies to prevent them from choosing Huawei in the construction of 5G.”
Healthcare	17	Oct1: “An important task of the new Alice is to help [elderly people] keep a daily structure.”
Regulation	51	Jan5: “Regulation and law will become important in the coming ten years”

The theme of climate change occurred 17 times, divided over 5 unique newspaper articles. Most of these articles discussed AI as a tool that might help to reduce climate change and it’s negative effects. For example, an article that discussed the use of AI in the design of building projects (Jun4) stated: “If you try to build sustainably you have to give up on other requirements, or it becomes less affordable. A computer can find that optimal balance, a person cannot”. Notably, none of the articles that mentioned climate change discussed the use and development of AI as contributing to the cause of climate change. The use of AI in healthcare was predominantly discussed as a positive application of AI as well. There were three unique newspaper articles that specifically focussed on the use of AI to find a cure for the Covid-19 virus or to prevent the spread of this virus. Ten other newspaper articles discussed other uses of AI in the area of healthcare. This included applications such as algorithms that help to diagnose various diseases, AI applications to support doctors during surgery, robots that help to take care of elderly people in nursing homes and the development of AI applications to be used in personalized medicine.

4.2.2.2 The impact of AI

A considerable amount of newspaper articles discussed what impact AI has on society, or what impact it might have in the future. A number of articles explicitly mentioned some of the philosophical concepts discussed in the theoretical framework, as is shown in table 9. However, the newspaper articles usually talked about more specific themes and types of impact. On the other hand, a few newspaper articles that described applications of AI mentioned there were ethical considerations to be taken into account, without specifying

what those ethical considerations might include or what impact the application might have on society. For example, one article (Sep1) described there was a lack of consideration of the impact hiring algorithms might have, stating: “Even though the technology helps to make the right decision, morally and ethically there is a gap”. In total, such unspecified implications were mentioned 23 times.

Table 9

Philosophical concepts in newspaper articles

Category	Frequency	Quotation
Autonomy	7	Feb1: “If it concerns matters of life and dead, it cannot be the case that the computer operates completely autonomously, according to the [European] Commission.”
Bias	22	Sep1: “People make a lot more mistakes than machines. Although biases are also ingrained in algorithms.”
Explainability	14	Aug1: “A problem is that even the creators of those algorithms don’t know exactly how they arrive at their translation or image qualification.”
Fairness	18	Mar3: “Too often the gain is for companies and the costs are for the society”
Responsibility	27	Oct5: “It is often unclear where the responsibility for “the product” begins and ends.”
Privacy	34	Jan5: “Meanwhile privacy disappears, as an expensive downside of the fact that everything appears to be free.”

As table 9 shows, of the concepts from the theoretical framework, privacy was mentioned most often in newspaper articles. As mentioned before, privacy was often discussed in relation to the European privacy regulation. In addition, privacy was regularly mentioned in articles about the use of AI for surveillance. There were three articles in which these topics co-occurred in the same sentence. For example, an article that compared the use of AI to prevent the spread of the Covid-19 virus in various countries (Mar2) stated: “In the USA politicians are proposing to increase the possibilities to collect private data in reaction to the virus”. This article continued by mentioning that stricter privacy laws prevent similar uses of AI in the European Union, which matches the earlier observation that the European privacy regulation was often mentioned in relation to comparisons between different areas of the world.

The concept that occurred least often in the newspaper articles is autonomy. A few newspaper articles mentioned the fear of AI becoming fully autonomous and overpowering people. However this was usually discussed as an example of a fear promoted in science fiction stories, instead of as a realistic risk. For example one article stated (Nov1): “There are many visions of fear surrounding AI: What if the computer autonomously develops itself into something we don’t want, like robots that see people as subordinate and destroy them?”. Similarly, multiple articles compared artificial intelligence to human intelligence

and concluded that AI cannot be autonomous and conscious like people. For example, an article (Dec1) argued: “Intelligence includes a lot more than carrying out calculation tasks. A computer that has consciousness? Unthinkable”.

On the other hand, there were a few articles that looked at how current applications of AI might already affect human autonomy. One article (Dec2) claimed: “The belief that we can make our choices largely autonomously is unfortunately rarely true”. This claim was followed by multiple examples of how we receive targeted information based on algorithmic predictions of social media platforms, Google, helpdesks and even supermarkets, which may limit people’s choices. Another issue that was mentioned a few times is that AI should not be allowed to make decisions that affect people autonomously without human control, as exemplified by the quote for autonomy in table 9. This issue relates to the question of who should be responsible for consequences of the use of AI systems, which was discussed more often.

In total, the issue of responsibility occurred 27 times in the newspaper articles. A few of the news articles focused specifically on the question of who should be responsible for AI systems, or argued that certain companies did not take enough responsibility. For example, in one article a professor was interviewed about this topic, and she stated: “We have to force companies to contribute to society, it is a question of taking responsibility”. The concept of “responsible AI” was mentioned multiple times as well, especially in news articles that discussed plans by governments and interest groups to invest in the responsible development of AI. In articles that focused on new applications of AI it was often stressed that the people using the AI application were responsible for checking the AI and making final decisions. For example, in an article about the use of AI in a military context, a general stated: “We apply a policy of meaningful human control, of human judgement”.

Even though the concept of fairness only occurred 18 times, there were many newspaper articles that discussed topics related to fairness. As exemplified by the quote for fairness in table 9, a considerable amount of the articles that discussed fairness were concerned with the increasing power difference between companies that create and use AI applications and the citizens that are affected by this. Similarly, other articles focused on power differences between governments and citizens. For example, an article discussing the European Union’s plans for AI (Mar1) stated: “UN-reporter Alston showed that new digital technologies deteriorate the interaction between governments and the most vulnerable people in society”. In total, the concept of power occurred 60 times and the concepts of fairness and power co-occurred in the same sentence six times. On a smaller scale, the concept of power was mentioned in relation to the interaction between programmers and machine learning algorithms as exemplified by the following quote (Aug1): “For normal people it remains hard to accept that you cannot follow a computer you programmed yourself anymore, that you can’t just turn a button if the computer confuses black people with gorilla’s”.

The aforementioned quote also points towards a discriminatory bias in an algorithm, which is an issue that was discussed in various newspaper articles. The concepts of bias and prejudice co-occurred in the same sentence 5 times and the concept of prejudice occurred 52 times in total. The concept of bias was often equated with prejudice, for example an article (Dec1) stated: “Another problem is “bias”, programmed prejudices”. However, there were a few articles that talked about bias in a broader sense. For example, in a newspaper article (Oct1) about healthcare robot Alice, a professor involved in the development of the robot explained: “Every algorithm represents a certain worldview. […] Imagine that Alice is going to find friends for you in a nursing home. Than it is important what the programmer thinks about friendship”.

An increase in the diversity of datasets used to train algorithms and among the people who make algorithms was seen as a possible solution for both types of bias. The concept of diversity occurred eleven times in total and five times in direct relation to discussions of bias and prejudice in algorithms. The articles in which diversity was mentioned in another context than as a solution for bias and prejudice in algorithms mainly discussed the importance of diversity in society and in specific groups, like government and higher education. For example, one article (Mar3) consisted of an interview with former politician Marietje Schaake, who said: “The Dutch “polder model” with diverse voices around a table is ideal for technological questions. Companies, interest groups, technical experts and governments each have a part to play”.

4.3 Results from the focus groups

4.3.1. Knowledge about AI

In the beginning of each focus group session the participants were asked about their first associations with the term "artificial intelligence". In all four focus groups the first participant to answer this question mentioned robots as one of their first associations with AI. In focus group 3, participant 13 did not talk about robots as a direct example of AI, but about robot films, saying: "My first association is apocalyptic robot films, in which robots become too smart and take over the world". Other examples of applications related to AI that were mentioned multiple times were smart devices and systems, virtual assistants, self-driving cars and Deepfake videos. Many participants also mentioned algorithms and machine learning, or self-learning systems or computers, as one of their first associations with AI. Some participants did not come up with the term "algorithm" by themselves, but could mention examples of the use of algorithms, like targeted advertisements and targeted content on social media.

Some participants also mentioned examples of technologies that are not clearly related to AI. For example, participant 1 mentioned satellites and Wi-Fi as technologies that he thought could maybe be artificial intelligence, saying: "That I find very special is that you can move whole pieces of text via Wi-Fi, like the text is going through the air". When talking about applications of AI in healthcare, participant 17 mentioned exoskeletons as a possible example, stating: "I had to think about exoskeletons for people with a spinal cord injury, that they can sort of walk again with the help of such a skeleton outside of their body". These are examples of technologies that many people find impressive, but in which AI typically does not play a role.

A few participants already started explaining what they thought AI was when asked about their first associations with AI. For example, participant 12 stated: "I see it in contradiction to what is not artificial intelligence, for example, normal statistics and calculations of averages and regression. And I think that with artificial intelligence you give the computer a lot more room to research what it wants to research, instead of giving it instructions yourself". In other focus groups the participants were asked to give a definition of AI. For example in focus group 2 the participants came up with a definition together, which included that computers and programmes are able to learn based on data they collect or that is provided by people, and based on what goes right and wrong. This is a considerably accurate explanation of machine learning.

4.3.2 Communication about AI

The participants were asked where they learned what they already knew about AI and what media they normally use to stay up to date about the news. Most participants said they had heard about AI through the news media they use. However, some participants also mentioned they had learned something about AI in their education. For example, participant 4 said: "I have at least heard the term in my education. That was in

the same range as virtual reality and things like that". A few participants mentioned they had seen some information about AI that was recommended to them online or by friends. Some participants also mentioned that they knew people who worked with AI, and had learned something about this topic through conversations with these friends.

All of the participants except for one said that they used an app to read news, the NOS app was mentioned most often as an example. Participant 15, who did not use a news app, explained that she didn't actively follow the news at all: "I usually only google things directly, for example with Corona I look for the RIVM [Dutch National Institute for Public Health and Environment] or things like that. I don't really follow specific apps or the news". Ten of the 18 participants said they sometimes read news articles that they found via social media, especially Facebook. After this, the most popular sources of news were news programmes and talk shows on TV, news websites, other websites and news programmes on the radio.

In each focus group the participants were asked to rank scientists, government, industry, traditional media and social media in order of who they trust most to who they trust the least. All participants said they trusted scientists the most, except for participant 11, who trusted traditional media the most, followed by scientists. She explained: "I hear and read less about scientists, so I know too little about scientists to know what to believe". All of the participant ranked their trust in social media the lowest, though sometimes this last place was shared with other stakeholders. Participant 5 explained she found it especially difficult to decide how much she trusted social media, because of the diversity of content on social media: "I find it hard to rank social media, because I follow a few platforms that I trust a lot on social media".

There was no clear agreement among the participants about how industry, government and traditional media were ranked in between the stakeholders that were trusted the most and the least. Some arguments that a few participants provided for putting traditional media lower on the list were that traditional media was more opinion based and that they tried to sell information by making stories seem more interesting. For example, participant 2 put traditional media in the middle of his list and explained: "Traditional media relatively often want to describe research results or other things in a way to make it more interesting". Participant 5 put traditional media on the second place on her list, followed by the government. She argued: "If there is someone that provides a critical view of the government every now and then, it is the media".

4.3.3 Impact of AI

Overall, the participants felt slightly more negative than positive towards AI. As table 10 shows, risks and fears of AI were discussed more often than hopes and benefits. However, most participants saw both risks and benefits of the different AI applications that were discussed. Notably, risks and benefits were even mentioned together in one sentence eleven times. For example, when the participants were asked to evaluate if their views about AI had changed after the focus group, participant 7 answered: "For me it didn't

change, it confirmed that AI can provide important benefits, but that we have to be very careful in what ways it is applied".

Table 10

Affective reactions to AI in focus groups

Category	Frequency	Quotation
Benefit	56	FG1 P1: "I have a Spotify account and if you make a playlist it gives you new suggestions and that way I get to know new songs. That's a fun advantage."
Risk	75	FG4 P17 (about deepfakes): "This is very dangerous, that in this way things can be published that someone has never said, even though it looks like they did."
Hope	12	FG3 P11: "I hope it can have a positive influence, for example in healthcare and in the climate change problem."
Fear	27	FG2 P5 (about deepfakes): "I find this pretty shocking. I realize that you can do a lot of damage to someone with this."

When talking about what AI might be able to do or not, the participants mentioned 19 affordances and 20 limitations of various AI applications. An affordance that was discussed in three of the four focus groups was that AI makes it possible for virtual assistants to recognize what you say and answer your question. For example, participant 2 stated: "I find it impressive that services like ok google or Siri can use, what I think is AI, to recognize what different voices say". As a limitation the need for further development of AI applications was mentioned a few times, especially for applications meant to reduce climate change. Participant 6 said about this: "Before it's really going to make a difference we still need to make a lot of steps to optimize it".

All of the philosophical concepts from the theoretical framework came up at least once during the focus groups. Table 11 shows how often each of the concepts were mentioned. The concept of privacy was discussed most often, it was mentioned at least three times in each focus group session. Like in the newspaper articles, privacy was frequently discussed in relation to surveillance, these concepts co-occurred in the same sentence five times. The discussions of privacy in relation to surveillance all occurred in the focus groups that discussed a newspaper article about the use of AI for facial recognition in surveillance. For example, in reaction to this article, participant 2 said: "If I have to hand something in for facial recognition or a similar technology, I find that a very small sacrifice if it means that a terrorist can be caught because of that". In focus group 4 the participants had a discussion about privacy in relation to targeted advertising. For example participant 18 worried that virtual assistants listen in on conversations to provide fitting advertisements, stating: "You also notice that if you talk to a friend about something and then a few hours later you suddenly see an advertisement about that". She also mentioned that she was not sure if virtual assistants actually listen to everything you say or if it was just an accident when this happens.

Table 11

Philosophical concepts mentioned in focus groups

Category	Frequency	Quotation
Autonomy	8	FG2, P7: "A lot of people are afraid that if machines get too smart it goes wrong, that it will transcend people."
Bias	9	FG1, P4: "They just start somewhere, I don't think they consciously left out [specific data in a training set]."
Explainability	7	FG3, P10: "The previous fragment said AI is a black box and you don't know exactly what happens, in this case that is possible and it is directly made insightful."
Fairness	4	FG3, P12: "I'm scared for what it can do with inequalities in society and the power of companies in society."
Responsibility	11	FG1, P4: "I think you always have to keep adjusting such a system. I don't think you can completely let it go at a specific moment."
Privacy	20	FG4, P16: "How privately can you still do things? I think not at all and sometimes I don't like that."

The second-most discussed concept during the focus group interviews was responsibility. The concept of responsibility only came up during the first and third focus group sessions. During focus group 3 responsibility first came up when the participants were asked to think of possible disadvantages of AI in the beginning of the focus group interview. Participant 10 reacted: "Who is responsible, because you cannot hold a computer responsible if something goes wrong". During the first focus group session responsibility first came up in reaction to the news fragment about discrimination in facial recognition algorithm, when participant 4 argued that people who make algorithms should take responsibility to make sure they train their algorithms with representative and diverse data. This discussion was continued in reaction to the article about the use of AI to detect cancer, which mentioned that there always has to be a doctor who checks the algorithm and makes a final decision. Participant 4 said about this: "I'm very happy that they don't completely rely on the AI algorithm, that there is always a final check by a doctor".

The idea that people should remain in control and should not rely on AI too much was important in discussions about autonomy as well. Again, participant four shared her fear that AI might learn itself something that people do not want it to learn, stating: "Maybe at a certain moment it becomes too smart and maybe it also decreases your own skills, because you might not look at it so critically anymore if the computer thinks for you anyway". This quote also exemplifies the sentiment that robots will probably not become autonomous in the sense that they will take over control like in science-fiction movies, but they might decrease peoples autonomy to the extent that they rely on AI to do specific tasks for them. This sentiment was shared by the majority of the participants.

Autonomy was also discussed in relation to the power individual people have to choose to participate in the use of AI technologies. For example, in response to a question

about the use of facial recognition algorithms in surveillance, participant 10 stated: “You don’t have free will in this situation, your face just gets recognized”. Participant 1 shared a similar opinion when asked how AI might impact his life, saying: “Whether I want it or not, it influences my life. We can’t live without it anymore”. This dependence on AI applications and the companies that create them was also discussed in relation to fairness, as the quote for fairness in table 11 exemplifies. The other three times that fairness was discussed it was in relation to possible misuse of facial recognition algorithms. For example, participant 11 said: “There could be difficult gaps or loopholes, that you could create fake evidence with a good mask or with Deepfake or by creating the face of someone else”.

The concept of bias was predominantly discussed in relation to the news fragment about discrimination in facial recognition algorithms that was used as an example in focus groups 1 and 3. For example, as a first reaction to this news fragment, participant 11 said: “I wanted to say that an advantage of AI is that it can’t be sexist or racist, but it turns out that’s not true, because it just depends on the source that the decisions are based on”. Outside of the discussions related to this news fragment, the concept of bias was mentioned only once, by participant 18 in focus group 4, who said: “If it is self-learning you don’t have control over what it learns and based on what types of factors it makes its choices”. Even though bias is not mentioned explicitly in this quote, it describes one of the problems that often underlies biases in algorithms.

This quote of participant 18 also hints at the issue of explainability of machine learning algorithms. The concept of explainability occurred seven times and in only one of the focus group sessions. Nevertheless, explainability was seen as an important issue in this focus group, participant 12 even included it in his definition of AI, stating: “[…] Thereby it’s important that you give the computer an assignment, then it figures it out in a way that you don’t understand and in hindsight you can understand what the computer did. There is a phase in which we cannot understand the computer”. In addition, participants in other focus group sessions did not mention explainability explicitly, but they did talk about the related topic of transparency. Participant 4 believed that the use of AI is deliberately untransparent in some situations, she explained: “It is artificial and it is intelligent, so it should not be too obviously present or annoy you. So I think they incorporate it in very insidious ways, so you don’t notice it as much”.

In relation to transparency, participant 18 mentioned the childcare benefit scandal as an example of a consequence of algorithms not being transparent. This was a political scandal in the Netherlands, in which thousands of parents were wrongly accused of making fraudulent benefit claims and which was partly caused by the use of an algorithm that predicted which parents were likely to commit fraud. Around the time the focus group sessions were organized, news media regularly reported about this scandal. Participant 18 referred to this scandal, stating: “I think if you look at the childcare benefit scandal, that it also has something to do with machine learning algorithms, and that it indeed gets out of control and it’s impossible to look at what the decisions are based on”. During another focus group session participant 5 referred to the same scandal when asked about a possible

disadvantage of the use of AI. She said: “I immediately thought about the childcare benefit scandal, the human dimension gets lost. If everything is decided by machines, where are the conversations and the personal contacts?”.

Some of the recurring topics from the expert interviews and media analysis were mentioned during the focus group sessions as well. Table 12 provides an overview of how often each of these topics were mentioned with examples of how they were discussed. The theme of healthcare recurred most often during the focus groups. In the first and third focus group session a news fragment about the use of AI to detect cancer was discussed. In both of the other focus group session the participants came up with applications of AI in healthcare themselves in the beginning, but it was also shortly mentioned in the fragment of the interview with Luciano Floridi that they discussed. In general, all participants were predominantly positive about the use of AI in healthcare, but they found it important that doctors remained responsible and critical when using AI.

Table 12

Recurring themes in focus groups

Category	Frequency	Quotation
Climate change	12	FG4 P18: “In the case of climate change I don’t think AI is the one and only solution, but I think it can help as support.”
Corona virus	7	FG2 P9: “We’re currently in the middle of the Corona crisis, […] I think we might see AI come back in a surprising way to get us out of this crisis, as one of the things that will contribute to that.”
Fake news	14	FG2 P8: “If it is said that something is fake I would personally belief that, but there are a lot of people who don’t believe it’s fake just because it says so, because who has added that and who decides that it’s fake?”
Games	2	FG1 P2: “I have seen that they use machine learning for simple games, that the programme learns by itself if I move in this way I lose, so I have to go the other way.”
Geopolitics	3	FG3 P12 (about facial recognition): “It’s not a good idea if you live in China and the government uses it to suppress the citizens.”
Healthcare	25	FG4 P15: “In the healthcare sector there aren’t a lot of applications yet, even though it could work very well, but we’re still a bit scared of that.”
Regulation	3	FG3 P14: “I hope there will be more rules and regulations for AI, I expect that there will be more attention for that and that limits will be set.”

The aforementioned interview with Luciano Floridi focused on applications of AI to reduce climate change, so most of the times climate change was mentioned, it was in relation to this news fragment. Even though the summary of the interview did not focus on this, some of the participants also mentioned the negative impact the development of AI can have on

climate change. For example, participant 9 said: “I saw a news article today that said that they are building a new windmill park near the Dutch coast and that half of the energy it will provide in the coming years has already been sold to Amazon. […] And not all of that goes to AI, but it is a part of the problem”. Other participants mentioned that having AI applications to help reduce climate change might discourage people from changing their behaviour in order to reduce their negative impact on the climate as well. Nevertheless, the participants were predominantly positive about the use of AI as an additional way to reduce climate change.

The concept of fake news was mentioned most often in relation to the news fragment about Deepfake video’s, which was discussed in focus groups 2 and 4, but it was referred to in the other focus group sessions as well. During the first focus group session participant 3 mentioned the spread of conspiracy theories via social media as a risk of AI, saying: “On YouTube it keeps linking you to the next video and then you might suddenly be watching a video about a conspiracy theory and maybe you start to believe it”. As the quote in table 12 exemplifies, several participants argued that the use of Deepfake should be regulated and that it should be made clear when a video is fake.

Apart from this, the concept of regulation was only mentioned three times in two of the focus group sessions, when the participants were asked about their expectation for AI in the coming five years. Participant 9 explained he expected that there would come more rules and regulations related to AI, stating: “Of course that always slightly lags behind the developments, but now it has been put on the agenda a bit more”. In relation to geopolitics, the participants only referred to the use of facial recognition for surveillance in China three times. The USA was mentioned a few times as well, but not in relation to geopolitical issues related to the use and development of AI.

A final recurring theme in the focus groups was that multiple participants came up with alternatives for the use of AI or with alternative technologies that had a similar impact. For example, in focus group 1 the participants compared the risk of losing certain skills by relying on AI applications to losing the skill to navigate by yourself when regularly using a navigation system. When participant 4 talked about the risk that doctors look less critically at AI applications meant to help them over time, participant 1 reacted: “That is the same as with navigation systems, we don’t think about it ourselves anymore”. In both focus group sessions that discussed the news fragment about Deepfakes, the participants came up with other techniques and technologies that could have similar effects. For example, when talking about the use of Deepfake to make actors look older or younger, participant 8 mentioned this could also be done using make-up. In focus group 4 participants compared Deepfakes to other things that make you doubt what is real and fake, like photoshop and advertising. Participant 15 said about this: “You currently see this a lot on social media, questions about what is actually real and fake”.

4.4 Comparison of results

4.4.1 Impact of AI

The experts, newspaper articles and participants of the focus groups all discussed more risks and fears than benefits and hopes related to AI. This difference was largest in the newspaper articles, in which risks and fears were mentioned 96 times whereas hopes and benefits were mentioned 50 times. In the focus groups this difference was smallest, with risks and fears occurring 75 times and hopes and benefits 68 times. When discussing specific applications of AI, the focus group participants mentioned 20 limitations, 19 affordances and no promises. The experts were slightly more positive about current applications of AI, mentioning 9 affordances, 5 promises and 6 limitations. The newspaper articles had an even more positive focus on AI applications, mentioning 97 affordances, 58 promises and 41 limitations. This is an indication that newspaper articles mainly reported positively about specific applications of AI, but negatively about AI in general. In the expert interviews and focus groups the discussions were more nuanced.

In the media analysis and focus groups the philosophical concepts from the theoretical framework that occurred most often were the same. The concept of privacy was mentioned most often, followed by responsibility and bias. During the expert interviews the concept of bias occurred most often, followed by explainability, responsibility and privacy. In both the expert interviews and focus groups the concept of fairness occurred the least often. However, issues related to fairness, such as prejudice and power differences, were among the most discussed topics in all three studies. Autonomy was the least discussed concept in the media analysis and the second-least discussed concept in the expert interviews. In the focus groups there was more attention for this issue, especially in relation to how reliance on AI might reduce people's autonomy. In the media analysis and expert interviews there was some attention for this issue too, but it was usually discussed in relation to responsibility.

There were a few other topics that reoccurred often. Firstly, the contribution of the development and use of AI to increasing anthropogenic climate change, which was discussed as a risk in the theoretical framework, reoccurred in the expert interviews and focus groups. However, in all three studies it was also mentioned that there could be ways to use AI in order to reduce the climate change problem. Secondly, topics related to the regulation of AI occurred in all three studies, this included discussions about existing laws and regulations, like the GDPR, as well as opinions on how AI should be regulated in the future. In relation to politics, in the media analysis especially, there was much attention for how the use and development of AI influences geopolitical relations.

Finally, healthcare was discussed as a promising area for the application of AI in all three studies. The possibility of AI being used to help cure people or possibly save their life was seen as a clear benefit. Applications in healthcare were often mentioned as one of the most impressive applications of AI in the expert interviews and focus groups. During the first focus group, participant 4 added the consideration that the expectation of privacy

people have in medical situations is already relatively low, so that is not a big drawback of the use of AI in this case. She explained: “In that case you’re already in the so-called medical mill, so you’re turned inside out anyway”.

4.4.2 Communication about AI

In the media analysis, most of the sources that were mentioned were affiliated with industry, followed by academia, governance and interest groups. Other media outlets, artists and citizens without clear affiliations with any of the aforementioned groups were mentioned a few times as well. In the expert interviews and focus groups most sources that were mentioned were categorized as media. This mainly included various news media, as well as specific books, films and tv-series. In the expert interviews, interest groups were the second most mentioned source. This was followed by sources affiliated with industry, academia and governance, which were all mentioned three times. In the focus groups sources affiliated with industry and governance occurred most often after media sources. During the focus groups the group of citizens was mentioned relatively often as well, since multiple participants said they talked about AI with people they knew.

During the expert interviews, the participants were asked about their opinions on how news media cover AI. Most participants were relatively content with how news media discuss AI, though multiple participants mentioned that news articles often lack nuance about what AI can and cannot do. In the focus groups, participants were asked to rank how much they trusted five different groups of stakeholders, including traditional news media. Most participants put traditional media on the second, third or fourth place in their ranking, which was often based on whether they perceived news media to be more or less independent from governments and private companies. Some participants of the focus groups also mentioned that news media often make stories sound more interesting, which matches with the experts’ observations that news articles sometimes lack nuance.

5. Discussion

In this chapter the main research questions and sub-questions will be addressed, starting with the research questions from a philosophical perspective, followed by the research questions related to communication science and ending with the main research question. This will lead into a discussion of the theoretical and practical implications of this research. Following this, the limitations of this research will be discussed and recommendations for further research will be provided. Finally, some main conclusions will be drawn.

5.1 Discussion of results

5.1.1 Impact of AI

5.1.1.1 Philosophical literature about the impact of AI

The first sub-question about the impact of artificial intelligence was: "What are the main considerations about the societal impact of AI in philosophical literature?". In the theoretical framework six main concepts were selected to categorize the topics that were most prominent in philosophical literature about the impact of AI. Those concepts were autonomy, responsibility, fairness, bias, explainability and risk. The final concept of risk included a variety of risks of the use and development of AI that were discussed in philosophical literature, but did not fall under any of the other concepts. This included risks of AI harming privacy and of AI harming the environment, through its contribution to causing climate change.

For the concept of autonomy, there are two main trends in philosophical discussions about the impact of AI. The first trend focusses on the question of to what extent AI can become autonomous and how similar artificial intelligence is to human intelligence (Helm et al., 2020; Johnson & Verdicchio, 2018). The second trend focusses on how AI might impact the autonomy of people who use or are affected by the use of AI (Hayes et al., 2020). Since current developments in AI are mainly directed towards making specialized AI systems that can outperform human experts in specific tasks, the discussion about the impact of the use of AI on people's autonomy is more relevant to the current impact of AI and the public debate. The philosophical discussions about autonomy are often related to discussions about responsibility in relation to AI as well.

Regarding responsibility, the main question is about who should be responsible for the consequences of the use of AI systems. Johnson and Verdicchio (2018) proposed a new type of agency, called triadic agency, that takes into account the artifact, the user and the designer that caused something to happen together. Triadic agency can be used to assign responsibility in cases where AI is used and it is not immediately clear who is responsible for the consequences (Johnson & Verdicchio, 2018). Hayes et al. (2020) argued that transparency about AI is important for responsibility, since information about what happened and who or what was involved is needed to hold someone accountable for the consequences.

Philosophical discussions about fairness and AI predominantly focus on how the use of AI might lead to unfair treatments of people. The High-Level Expert Group on Artificial Intelligence (AI HLEG, 2019b) distinguished between substantive fairness, which entails that AI systems should ensure that benefits and costs are distributed equally and justly and prevent unfair biases, and procedural fairness, which means that it is possible for people to contest and effectively rectify decisions made with AI systems. Most philosophical discussions about fairness and AI seem to focus on substantive fairness, mentioning that algorithms can inherit biases from data and AI developers, which can lead to unfair results (Binns, 2018; Hayes et al., 2020).

Even though the concept of bias was often related to unfairness in algorithms, there were philosophical discussions about other types of biases in AI as well. Kitchin (2017) argued that algorithms necessarily include some forms of bias because they categorize and sort data and because they are developed and used in relation to a specific context. Dobbe et al. (2018) further distinguished between three types of bias that may arise in different stages of the development and use in AI. These three types of bias are pre-existing bias, technical bias and emergent bias (Dobbe et al., 2018, p. 1).

The concept of explainability is related to the opacity of AI systems and machine learning applications in particular. Burrell (2016) argued that machine learning algorithms that are used to classify information are usually opaque in the sense that it is unclear how or why the inputs of an algorithm lead to certain decisions. This opacity can be a problem when deciding whether an AI system treats people fairly and when assigning responsibility for consequences that were partly or completely caused by AI. Therefore, the AI HLEG (2019b) argued that AI systems need to be explainable to those affected by them as far as possible and that development processes of AI need to be transparent. In addition, Hayes et al. (2020) argued that it should be possible to get knowledge about AI that is accessible and explainable.

As mentioned before, the concept of risk was added to include various other risks of the use and development of AI that are regularly discussed in philosophical literature about the impact of AI. Privacy was discussed as the first issue that received much attention (AI HLEG, 2019b; Raab, 2020). AI applications might threaten privacy through how they create and categorize groups and through the movement of personal data between contexts (Hayes et al., 2020). The second risk that was discussed was that AI might play a role in causing or contributing to harm done to people and the environment. Discussions of how the development and use of AI contributes to climate change by Ensmenger (2018) and Strubell et al. (2020) were highlighted. Two other risks that were regularly mentioned, but not discussed in-depth were risks related to the safety and accuracy of AI and risks related to ownership and property.

5.1.1.2 Expert's considerations about the impact of AI

The second sub-question about the impact of AI was: "What are the main considerations about the societal impact of AI among experts in the field?". In total, all of the concepts

that were selected from the philosophical debate in the literature review, were mentioned at least once during the expert interviews. The participants of the expert interviews mentioned the concepts of bias, explainability, responsibility and privacy most often. The concepts of autonomy and fairness were only mentioned by one participant, who also had a background in philosophy. The risk of AI contributing to the climate change problem was mentioned as well, though a few participants explained that AI might also be used to help reduce climate change. Other topics that emerged in the expert interviews included the role of AI in distributing information, for example through targeted advertising and recommendation systems, and laws, regulations and political issues surrounding AI.

Even though the concept of fairness was mentioned only once, topics related to this concept, like discrimination, prejudice, diversity and power differences between people and organizations recurred often during the expert interviews. These discussions were often related to the concept of bias as well, since biases in algorithms and training data can lead to increased power differences. Some of the participants also explained that biases are inherent to any machine learning algorithm, because data need to be sorted and prioritized in some way to get a useful output. Yet even if a machine learning algorithm has no unfair biases and a very small error margin, machine learning algorithms still have a risk of leading to large negative consequences, due to their scalability. This scalability was mentioned as an important aspect of machine learning by multiple participants. It means that once a machine learning algorithm has been developed it can be applied on a very large scare. One of the participants argued that in addition to this problem, the people for whom a mistake is made, currently have hardly any options to rectify that mistake.

This problem is related to the concept of explainability. Even though the technology and the mathematics behind machine learning algorithms are not extremely complicated, as multiple participants of the expert interviews emphasized, it can still be difficult to explain to people why the algorithm provided a certain result. This is complicated further by the fact that these algorithms are often developed and used by large, powerful companies. As Burrell (2016) argued, these companies may deliberately keep their algorithm opaque. Even if this is not the case, it is hard for individual people that may be harmed by algorithms to hold these companies accountable. Notably, the concepts of autonomy, responsibility, fairness, explainability and bias all emerge in this example of the risks that the scalability of machine learning algorithms may pose.

5.1.1.3 Public debate about the impact of AI

The third sub-question about the impact of AI was: “What considerations about the societal impact of AI are apparent in the public debate about AI?”. The media analysis and focus groups provided insights to answer this question. All of the selected philosophical concepts were mentioned at least a few times in the newspaper articles that were analysed and in the focus groups. In both the media analysis and the focus groups, the concepts of privacy, responsibility and bias occurred most often. Privacy was often discussed in relation to the use of AI in surveillance or in relation to regulations and politics. Especially in the

newspaper articles, the concept of bias was regularly conflated with the inclusion of prejudices in algorithms.

In the focus groups these topics were often linked to each other as well. In relation to responsibility, the focus group participants emphasized that it was important to them that human experts remain in control when they use AI applications. The consensus was that human experts should always be able to check the outcome of AI applications and have the final responsibility. This idea recurred in newspaper articles as well. In addition, newspaper articles paid more attention to whether AI developers and companies currently take enough responsibility for the systems they create and implement.

Like in the expert interviews, the participants of the focus groups mentioned how AI could contribute to causing climate change and to possibly solving this problem. Interestingly, the newspaper articles mainly focused on the positive contribution AI could have in helping to diminish climate change and did not pay much attention to the negative impact. The topics of the role of AI in distributing information and the regulation of AI, that emerged in the expert interviews, recurred in the public debate as well. In the media analysis especially, there was a lot of attention for geopolitical tensions related to the development of AI, some articles referred to this as a race to become a world leader in AI. In the focus groups there was less attention for these geopolitical issues and more attention for current and future regulations of AI. In relation to the role of AI in distributing information, algorithms that recommend advertisements or content on social media and streaming platforms were among the examples of AI that recurred most often in the focus groups. In the newspaper articles that were analysed these types of algorithms were discussed regularly as well.

5.1.1.4 Alignment of the philosophical, expert and public debate about the impact of AI

The main research question related to the impact of AI was: *How well aligned are philosophical discussions of AI with expert, media and public views and what consequences do current misalignments have for both philosophy and science-society relations?* The concepts of autonomy, responsibility, fairness, bias, explainability and risk, which are important in philosophical literature about the impact of AI, were mentioned in the expert interviews, media analysis and focus groups as well. This shows that the philosophical, expert and public debate are relatively well-aligned with each other. However, there are some misalignments in which concepts receive most attention and how they are discussed.

The first misalignment concerns the concept of autonomy. Autonomy received a lot of attention in the philosophical discussion about the impact of AI, but not as much in the expert and public debate. As mentioned before, there are two main trends in discussions of autonomy and AI in philosophical literature. The first trend focusses on the question of to what extent AI can become autonomous. In the media analysis this was mainly discussed as a topic in science fiction, art and games about AI. In addition, there was a very small amount of newspaper articles about new developments in AI that shortly mentioned as a sidenote that people did not need to worry about AI becoming fully autonomous and

overthrowing people. In the focus groups the fear of AI taking over the world was mentioned as an unrealistic fear or science fiction scenario a few times as well. However, a few participants also expressed some genuine worries about how far the self-learning aspect and autonomy of AI could go and if it would still be possible to correct and control AI if it develops by itself in a direction the people who created it did not expect. Some of the participants in the expert interviews also mentioned they thought this was an understandable worry for people without expertise in AI to have.

The second trend in the philosophical debate about autonomy and AI concerns how AI might impact the autonomy of people who use AI and of people who are affected by the use of AI. Hayes et al. (2020) explained this by focusing on how decision making algorithms affect the autonomy of both the decision makers and the people who are subject to the decisions that are made. In the focus groups there were a few discussions about how AI might impact the autonomy of those who use AI. The main worry in these discussions was that experts, like doctors using algorithms to detect cancer, would rely too much on AI, which could lead to a decrease in their own ability to detect cancer and to be critical of the results of the algorithm they use. This issue of deskilling has received some attention in the philosophical debate, though not in relation to the impact of AI on autonomy (Carter et al., 2020; Cowls & Floridi, 2018).

The question of how AI might impact the people who use it did not recur clearly in the media analysis or in the expert interviews. The impact that AI might have on the autonomy of people that are subject to decisions made by AI was mentioned explicitly by one participant of the expert interviews. Another participant did not mention autonomy, but talked about how most people are powerless if a decision made by an algorithm has a negative impact on their life, which shows they have little autonomy in such a situation. The large power difference between companies that develop AI and the people who are subject to decisions made by AI was discussed in the analysed newspaper articles and in the focus groups as well.

Instead of focussing on the fear of fully autonomous robots in science-fiction scenarios, news media should pay more attention to how AI might impact the autonomy of those using AI and those affected by this use. In doing so, they could provide the public with a more accurate perspective on how AI might impact their life currently and in the near future. The inclusion of autonomy in the public debate could help make it easier to understand the impact of the increasing power differences between large technology companies and citizens. In addition, it can provide a better understanding of what is at stake when AI is used to make decisions that have a large impact on people's lives, such as in loan allocations, hiring procedures and the judiciary system.

The second misalignment is related to the concept of privacy. Privacy was the most mentioned concept in the media analysis and the focus groups, but received less in-depth attention in the philosophical debate about AI. In the expert and public debates, the topic of privacy was often discussed in relation to laws and regulations, like the GDPR. When the participants in the expert interviews were asked if there were any procedures to draw

attention to ethical issues in their work, privacy regulations came up most often as well. The attention to privacy among experts and in the public debate is good, but it might also draw attention away from other important topics related to the impact of AI. For example, giving consumers more control over their privacy increases their degree of autonomy to a certain extent, but there are other aspects of AI that limit consumers autonomy, like the lack of explainability, that receive less attention and remain unsolved.

In the philosophical debate about the impact of AI, privacy has received less attention than in the public debate. Even though privacy protection is one of the most mentioned issues in ethical guidelines for AI (Raab, 2020, p. 4), philosophical articles that provide an in-depth analysis of how AI might impact privacy are rare. This may partly be related to the existing regulations about privacy that exist already, which show that there are ways to solve at least some issues related to privacy outside of the scope of philosophy. Another reason may be that privacy has already received a lot of attention in philosophical debates about other technologies, like the internet and other information technologies. This is understandable, but it would be interesting to see philosophical considerations that focus on the impact of AI on privacy specifically. In this way it could be explored if there is anything inherently different about AI that might lead to new risks or opportunities in relation to privacy.

The third misalignment is about how the concept of bias is discussed. In the public debate bias was almost exclusively mentioned in relation to prejudice and discrimination in algorithms. In the philosophical literature and expert interviews there was attention for how biases in AI applications can lead to the unfair treatment of people as well. However, other types of bias were discussed as well. Since AI systems analyse and sort data, they inherently include biases. These biases do not necessarily cause discriminatory decisions or unfair treatments of (minority) groups or people. Nevertheless, as one of the participants in the expert interviews explained, any type of bias, even the necessary ones, can have a harmful impact if AI is applied on a very large scale. This problem is currently overlooked in news media.

The fourth misalignment relates to the large scale of application of AI and machine learning applications. This issue arose in the expert and public debate, but does not seem to receive as much attention in philosophical literature about the impact of AI. In the expert interviews, scalability was seen as one of the most important aspects of machine learning. In the media analysis and focus groups scalability was not mentioned explicitly, but there was much attention for how the development and use of AI affects power differences between people and societies. In the media analysis, the impact of AI on geopolitical relations recurred often, with newspaper articles talking about an AI race between countries and focussing on how much money different governments were spending on the development AI. In the focus groups power differences between governments and citizens or private companies and citizens were discussed more often. These topics also recurred in the expert interviews and media analysis.

As mentioned earlier in the discussion about experts considerations about the impact of AI, the issue of scalability relates to the concepts of autonomy, responsibility, fairness, explainability and bias as well. Some philosophical articles also mentioned the scale of application of AI as part of their discussion of other topics. For example, Burrell (2016) discussed how the large scale of application of machine learning algorithms makes them more opaque and less explainable. Nevertheless, the scalability of AI and its impact on power relations currently does not receive much attention as a separate issue in the philosophical debate about the impact of AI. Since this is one of the impacts of AI that experts and members of the public are most worried about, it would be helpful if philosopher's paid special attention to the scalability of AI applications and what that means for the impact they have on the world.

The final misalignment is about the differences in where the main focus lies in the philosophical, expert and public debate about the impact of AI. Firstly, there is an overwhelming negative focus in the philosophical debate about AI. Even though risks and fears of AI were discussed more often than hopes and benefits in the three empirical studies, there was usually some weighing of costs and benefits. In contradiction, the philosophical debate focuses almost exclusively on risks. Even if philosophers argue that AI should be used more in a specific situation, they often talk about this as a risk of the underuse of AI (Cowls & Floridi, 2018; Hayes et al., 2020; Kitto & Knight, 2019). There are some exceptions to this negative focus, for example, a philosophical article by Floridi et al. (2018) weighs risks and benefits of AI and provides recommendations to take into account when implementing AI. Nevertheless, the predominant focus on risks in the philosophical debate about AI, may also be part of the cause of underuse of AI.

A possible cause for this negative focus in philosophical discussions about AI is that most discussions about the impact of AI focus on abstract concepts. The six philosophical concepts of autonomy, responsibility, fairness, bias, explainability and risk exemplify this. These concepts are often derived from basic human rights or broad philosophical theories. However, in order to improve public understanding of AI and its impact, the philosophical debate should include more practically oriented philosophical articles that weigh risks and benefits of applying AI in specific situations. For example, philosophical frameworks based on utilitarian ethics could be used for this. The focus on the application of AI in specific contexts is important too, since risks and benefits of AI are likely to be different for different areas of application if AI. It would be helpful to have philosophical considerations about whether or not certain benefits may outweigh certain risks in specific situations. These could serve as examples for similar situations and could be used by science communicators to increase public awareness of how AI might impact citizens in various situations.

5.1.2 Communication about AI

5.1.2.1 The views and expectations of experts in AI

The first sub-question about the communication about AI was: "What views and expectations do experts in the field of AI have about artificial intelligence?" During the

expert interviews it was confirmed that there is no clear consensus about specific definitions of AI and machine learning. Multiple participants of the expert interviews mentioned they found it difficult to give a short, coherent definition. Nevertheless, there were some characteristics that recurred in most of the explanations that the participants provided. Most participants saw the term "artificial intelligence" as an umbrella term for various technologies that can analyse their environment and take certain actions based on that analysis. Machine learning is one of those technologies, for which the participants mentioned the self-learning aspect, pattern recognition and the scalability of applications as important characteristics.

The participants of the expert interviews mentioned both risks and benefits of the technologies they worked with and AI in general. All participants stated that they felt responsible to think about the possible impact of their work with AI and to prevent any possibly harmful effects where possible. A majority of the participants of the expert interviews thought they were more aware of negative impacts that AI might have on society than most people around them. Most of the participants also felt responsible to inform others about AI and its impact, as part of their job or in their free time. This included informing colleagues, teaching students and informing publics by writing articles, giving public talks and joining events.

All of the participants of the expert interviews voiced some positive expectations for future developments and applications of AI. The main expectations for the development of AI were that it would become more efficient and would be able to do more complicated tasks, because of the availability of funding and rapid increases in computing power. Expectations related to the application of AI included that AI would be applied a lot more often in many different organizations, including in smaller companies, since AI applications are becoming more accessible and easier to use. Finally, a few of the participants expected that more laws and regulations for the development and use of AI will be made and enforced in the coming years.

5.1.2.2. The representation of AI in news media

The second sub-questions related to communication about AI was: "How do news media report about artificial intelligence?". The first topic to be discussed regarding this question is how news media, in this case newspaper articles, frame the topic of artificial intelligence. In the in-depth media analysis of 53 newspaper articles about AI from the main Dutch newspapers, impact frames, issue frames and risk and benefit frames were distinguished. Most newspaper articles used societal impact framing, followed by group impact and personal impact framing. This is similar to the results of a study on the study by Chuan et al. (2019) about how AI is framed in American newspapers. They distinguished societal impact, personal impact and mixed framing and found that societal impact frames occurred most and personal impact framing occurred least in their sample (Chuan et al., 2019). Chuan et al. (2019) also found that the topics of threat, politics/policy and ethics were often discussed with societal impact frames. This is similar to the observation that the news

articles in the media analysis that used impact frames often discussed the impact of AI on society and citizens in general, without focussing on a specific application of AI.

In the media analysis, a small majority of 30 articles was framed around episodic issues, which means that they mainly focused on a singular incident. Again, this corresponds with the findings of Chuan et al. (2019) that the majority of the articles they analysed were framed around episodic issues. Chuan et al. (2019) also found that newspaper articles that discussed topics related to business and economy were more likely to use episodic framing, whereas articles about topics related to threats, politics and policy were more likely to use thematic framing. This partly matches with the observation that most of the articles about episodic issues in the media analysis discussed new applications and breakthroughs in the development of AI, whilst the articles about thematic issues often focused on the impact of AI and larger trends in the development of AI. However, the media analysis also included multiple articles that discussed changes in funding and regulations of AI, which relate to the topics of politics and policy, but used episodic issue framing.

Overall, risks were used a lot more often to frame AI in the newspaper articles than benefits. In addition, fears of AI were discussed more often than hopes for AI. This contradicts the findings of Chuan et al. (2019) that in American newspaper articles about AI benefits were discussed more often than risks. This can partly be explained by the fact that some of the topics that Chuan et al. (2019) coded as risks and benefits were coded as affordances, promises and limitations of AI applications in this study. Affordances and promises of AI were mentioned a lot more often than limitations of AI. This matches with the observation that news articles about specific applications of AI often had a more positive focus than articles that discussed the influence and societal impact of AI in general. When the affordances, benefits, hopes and promises are taken together these positive assessments of AI occurred more often than the risks, limitations and fears.

The second aspect related to how news media report about AI was which sources are mentioned most often in newspaper articles. Members of all four groups of the quadruple helix were regularly mentioned as sources in the articles included in the in-depth media analysis. Sources affiliated with industry were mentioned most often, followed by individuals and organizations associated with academia and governance. Citizens were only mentioned 7 times, but interest groups, which often represent citizens were mentioned a lot more often. This corresponds with the observation of Siune et al. (2009) that citizens are usually only actively involved in science as members of other stakeholder groups.

5.1.2.3 The views and expectations of laypeople about AI

The final sub question related to communication about AI was: “What knowledge, views and expectations do laypeople have about artificial intelligence?”. Most of the participants in the focus groups could come up with accurate explanations and examples of AI. Multiple participants correctly associated AI with self-learning systems, robots and algorithms in general or specific types of algorithms. Many of the participants who were still studying or had recently completed studying at a university learned something about AI as part of their

education, even if their education was not related to computer science or engineering. The rest of the participants mainly relied on news media and conversations with people they knew to get information about AI. Some participants also mentioned that they mainly learned about the opinions of other stakeholders, like scientists and politicians, through news media as well. Even though none of the participants mentioned that they read physical newspapers, many participants mentioned they regularly read news articles online or consumed news programmes on the radio and tv.

This reliance on news media corresponds with the large overlap in recurring topics in the media analysis and focus groups. The three philosophical concepts that recurred most often were the same for the media analysis and the focus groups and were discussed from a similar point of view. For example, privacy was usually discussed in relation to regulations or the use of AI in surveillance and bias was often related to algorithmic prejudice and discrimination. Nevertheless, the participants of the focus groups had a more nuanced view of AI than the newspaper articles. The newspaper articles usually described specific applications of AI in a positive way, but the impact of AI in general in a negative way. In the focus groups, risks and fears only occurred slightly more often than hopes and benefits related to AI and the amount of affordances and limitations of AI that were mentioned were almost the same. The observation that focus groups provide more nuanced views is in line with earlier research that showed that the interaction between participants in focus groups can help to uncover nuances and complexities (Cyr, 2016, p. 248)

The participants of the focus groups were asked to rank how much they trust scientists, government, industry, traditional media and social media. The majority of the participants said they trusted scientists the most and social media the least. The high trust in science among the participants is in line with previous research on trust in science among Dutch citizens (Broek-van den Honingh & de Jonge, 2018). Broek-van den Honingh and de Jonge (2018, p. 11) asked participants to rate their trust in science, the judicial system, trade unions, newspapers, television, the government and large companies. They found that, on average, the participants trusted science the most and large companies the least (Broek-van den Honingh & de Jonge, 2018, p. 11). Newspapers and television ended up in the middle of the ranking of how much trust participants had in the different institutions (Broek-van den Honingh & de Jonge, 2018). This corresponds with the results of the focus group interviews, in which the majority of the participants placed traditional media somewhere in the middle of their ranking list.

5.1.2.4 The relation between expert, media and public views of AI

The answers to the three sub-questions related to communication about AI can be used to address the main research question: “How do views and expectations about AI discussed by experts, news media and publics relate to each other and what insight does this give for understanding the science-society relationship?” First of all, the participants of the expert interviews, the newspaper articles that were analysed and the focus group participants all

focused more on the risks than on the benefits of AI when talking about AI in general. This focus was even more noticeable in the philosophical literature about the impact of AI, which almost exclusively discussed risks, whereas the expert and public debates usually weighed risks and benefits against each other. When specific applications of AI were discussed, the experts and newspaper articles mentioned affordances and promises more often than limitations. In the focus groups this division was more equal, with 20 mentions of limitations and 19 mentions of affordances of AI applications.

Secondly, there was a clear overlap between the topics that received the most attention in the newspaper articles and in the focus groups. The philosophical concepts that recurred most often were the same in the media analysis and in the focus groups and other topics, like healthcare, climate change, discrimination and power differences recurred in both studies as well. Nevertheless, the focus group participants often had more nuanced views on these topics, whereas news articles tended to focus on either the positive or the negative aspects. The participants in the expert interviews also had a more nuanced view than the news articles provided, though they often focused on slightly different aspects in their discussions of AI. Overall, most participants of the expert interviews were relatively satisfied with how news media report about AI. Their main point of criticism was the lack of nuance in news articles. Some participants mentioned that news articles often provided a positive view of specific AI applications and a negative view of the impact of AI in general. This was confirmed by the media analysis and by earlier research from Chuan et al. (2019).

Thirdly, the media analysis showed that all four groups of the quadruple helix were represented in news media about AI. Sources affiliated with industry, academia and politics were mentioned most often as sources in newspaper articles about AI. Citizens were represented as well, though mainly through interest groups. The participants in the expert interviews and the focus groups often mentioned they read news articles to learn about new developments in AI. As can be expected, for the focus group participants news media were usually their main source of information about AI, whereas the participants in the expert interviews also mentioned scientific journals, industry publications and events as important sources of information.

Even though most participants of the focus groups relied on news media to receive information about developments in science and technology, they did not always have a high degree of trust in news media. With one exception, all participants had most trust in scientists out of the stakeholders of science, industry, government, traditional media and social media. Reasons for trusting traditional media that the participants mentioned included that traditional media aim to inform people and that they provide a critical perspective. The participants who had less trust in traditional media explained they thought news media made stories sound more interesting or focused too much on opinions in order to sell information.

The level of trust that citizens have in science and media is an important element of the science-society relationship. As discussed earlier, recent developments in how media

and individuals share information have posed new threats for the public trust in media and science (Scheufele & Krause, 2019). One of these threats is the increase in the spread of misinformation about science and how this may impact the publics' trust in news reports about science (Scheufele & Krause, 2019). The results of the focus groups showed that trust in science among Dutch citizens is high, but their trust in media is lower. This corresponds with the findings from Broek-van den Honingh and de Jonge (2018, p. 11), who showed that science was the most trusted institution in in all three of their monitors of 2012, 2015 and 2018. They also showed that trust in newspapers and television has slightly decreased over time, according to the three monitors (Broek-van den Honingh & de Jonge, 2018, p. 11).

The results of this research about the case of AI can provide some insights that can be used to increase understanding of the science-society relationship. The participants of the focus groups gave a few reasons for having less trust in traditional media than in other sources. The most important reasons were that they thought that news media focused too much on opinions and that they made stories sound more interesting. In addition, the main critique the participants of the expert interviews had about news reports about AI was that they lacked nuance. Both results point towards the possibility that describing AI in more nuanced ways could increase publics' trust in news articles about AI. This could help to make communication about AI more effective and improve science-society relations.

On the other hand, one participant of the focus groups mentioned that a reason for her to have much trust in traditional media is that they provide a critical view on the government. This corresponds with the argument made by Scheufele (2014) that the socio-political context in which science communication occurs should be taken into account, since it influences how effective science communication is. In this case, how news media report about politics might have an impact on how much trust publics have in news reports about AI as well. Thus, taking the context in which publics encounter science communication into account, can increase the understanding of the science-society relationship.

5.1.3 Science-society relationship

The overarching research question for this thesis was: *What insight does the case of AI give on the role of philosophy and communication in increasing understanding of the science-society relationship?* Communication and philosophy both have an important role to play in increasing the public understanding of AI. Bringing together insights from the philosophical debate and the public debate about the impact of AI adds value to both discussions. If philosophical, expert and public debates about AI and its impact on society are well-aligned, members of society can get a good understanding of how AI might impact their life.

Combining literature analyses of academic publications in philosophy and communication science with three empirical studies provided a holistic overview of how AI might impact individual citizens and society at large. The combination of the analysis of

scientific literature and the expert interviews resulted in a rich understanding of the current state of development of AI and the most important discussions and expectations in the field. Taken together, the large-scale and in-depth media analyses of newspaper articles and the focus groups provided a good representation of the public debate about AI and its impact. The results of the media analysis were used to guide the focus groups, with news articles from the sample of the media analysis being selected as examples of applications of AI. This mirrors how discussions in the public domain are based on how news media report about AI.

The focus on the philosophical concepts of autonomy, responsibility, fairness, bias, explainability, and risk in the three empirical studies provided an in-depth view that highlighted some specific nuances in the public debate about AI. It showed what areas of application and what types of impact of AI currently receive most attention. Notably, in the expert interviews, media analysis and focus groups healthcare and climate change were seen as areas in which AI could contribute to positive solutions. On the other hand, questions about responsibility for AI, the contribution of AI to discrimination and the impact of AI on power differences and privacy recurred as important problems. It also highlighted some topics that should receive some more attention in order to optimally inform citizens about how AI might impact their life, like discussions of bias in algorithms that are not directly related to prejudice and the impact AI can have on the autonomy of users and subjects of AI applications.

The analysis of how AI was framed in newspaper articles also provided new insights for the philosophical debate about the impact of AI. Firstly, the expert interviews, media analysis and focus groups brought out some topics, especially related to privacy, power differences and the scalability of AI, that deserve more attention in the philosophical debate. In addition, the focus on risk and benefit frames brought out some new insights about the philosophical debate about AI. Even though the difference between how often risks and benefits were mentioned was largest in the media analysis when comparing the three empirical studies, it also became clear that the focus on risks is even more apparent in philosophical literature. Philosophical articles about the impact of AI focused almost exclusively on risks, whereas in the public debate risks and benefits were usually weighed against each other.

This focus on risks in the philosophical debate may also lead to underuse of AI. People might choose not to use AI even though it would be beneficial, because they fear it will cause too much risks. To prevent underuse, the focus in the philosophical debate should shift from discussing risks of AI related to abstract philosophical concepts to the inclusion of more practical discussions about the impact of applications of AI in different contexts. For example, moral frameworks based on utilitarian ethics could be used to provide in-depth, philosophical analyses that weigh risks and benefits of specific uses of AI against each other. In turn, these practically oriented philosophical discussions could serve as examples for science communicators to increase public awareness of how AI might impact citizens in various situations.

5.2 Theoretical implications

5.2.1 Theoretical implications for philosophy

The impact of AI on society and on individual citizens has already received much attention in philosophical theory. In this area there is much interdisciplinary collaboration between experts in philosophy, law, social science, computer science and engineering as well. This research added a public perspective to the various views that are already represented in the philosophical debate. The debate about new technologies in the public domain often differs from that in the philosophical domain. This has been shown in earlier studies about different topics, for example in a review study by Dijkstra and Schuijff (2016) for the topic of human enhancement. Since the moral frameworks that philosophers use are limited, the philosophical discussion about AI might miss some issues that citizens worry about. The empirical studies in this research helped to show the extent to which the issues that philosophers deem most relevant align with what experts and publics see as the most relevant issues related to the impact of AI.

This research has shown that in relation to the impact of AI, there is some overlap between the philosophical and public debates, but the emphasis is put on different topics. Including the new perspectives from newspaper articles and citizens without expertise in AI can enrich the philosophical debate about the impact of AI. Based on the results of the expert interviews, media analysis and focus groups, there are a few topics that should receive more attention in the philosophical debate. Firstly, the impact of AI on privacy and on power differences were among the risks that recurred most often in the public debate, but have received little in-depth consideration in philosophical literature about the impact of AI. Secondly, there is a need for philosophical deliberation on the scalability of AI applications and machine learning in particular. This would help to evaluate if the large scale of application of AI is of special importance and if it possibly leads to new impacts which have not been considered before.

Finally, philosophical literature focuses almost exclusively on risks of AI, without explicitly weighing risks and benefits. This disproportionate focus on risks may contribute to an underuse of safe and helpful applications of AI. In addition, the philosophical discussion mainly focuses on risks related to basic human rights and abstract philosophical concepts, like the ones discussed in this research. These discussions are usually not directly related to the impact of specific AI implications applied in context. It is important that there remains room in the philosophical debate about the impact of AI to focus on abstract concepts that are not directly relevant for the public debate. However, it would be helpful if there was attention for more practical issues and the balance between risks and benefits in specific situations in which AI is applied as well. For example, philosophical frameworks derived from utilitarian ethical theories could be applied to the impact of AI, to weigh the risks and benefits of AI being applied in specific situations.

5.2.2 Theoretical implications for communication science

At the time of starting this research, communication about AI had not received much attention in scientific literature about communication science. A few media analysis studies had been conducted about how news media report about AI (e.g. Brennen et al., 2018; Chuan et al., 2019). This research compared a media analysis about how newspaper articles report about AI with an analysis of philosophical literature, expert interviews and focus groups with citizens, to provide a broader view on how media reports about AI are perceived. This can help to create a better understanding of differences between scientific and philosophical discussions on the one hand and discussions in the public domain on the other hand. It can also help to create a better understanding of the role of the media in bringing awareness of scientific findings to the public domain.

In addition, focussing on how the philosophical concepts of autonomy, responsibility, fairness, bias, explainability and risk recurred in discussions about the impact of AI in the public domain, provided a more in-depth analysis of these discussions. It brought out nuances in how the impact of AI is discussed in the public domain that would likely have been overlooked if these topics were not explicitly included in the analysis of the newspaper articles and focus groups. For example, instead of focussing only on how often different themes and topics recurred in each of the three studies, differences in how experts, news articles and citizens discussed the selected philosophical concepts emerged in the analyses as well.

Finally, the comparison of expert opinions on how news media report about AI with how much trust laypeople have in different stakeholders involved with AI provided some insights that can increase understanding of the science-society relationship. Both experts and laypeople pointed towards the importance of nuance in newspaper articles about AI. Currently, newspaper articles about specific applications of AI have a predominantly positive focus, whereas articles about the impact of AI in general tend to focus on negative aspects. By adding more nuance when discussing AI in general and specific applications of AI, there is more consistency between news articles, which could help to improve the efficiency of science communication about AI. This could also have a positive impact on how much trust publics have in science and news media in general.

5.3 Practical implications

In addition to the theoretical implications, this research also has some implications for the practice of science communication. Specifically, it offers some suggestions to improve the communication about AI and its impact in news media. First of all, in comparison to the expert interviews and the focus group discussions, newspaper articles were less nuanced in their discussions about AI. Articles about specific applications of AI often focused mostly on the affordances and positive aspects of these applications, without mentioning possible negative effects the use of these applications might have. On the other hand, news articles that discussed the impact of AI in general usually focussed mainly on possible negative impact, without mentioning positive aspects of AI. Adding more nuance in both types of articles would make it easier for people who read those articles what AI entails and how different applications of AI might impact them.

Secondly, there were a few topics related to the impact of AI that could be represented more accurately in news media. A few participants in the expert interviews mentioned that they thought news articles often made new applications of AI seem more impressive than they actually are. Both in terms of how accurate the results of AI applications are and in terms of how complicated they are. Even though algorithms are complicated, the basics of AI and machine learning can be explained in a way that is easy to understand for most people.

In relation to the philosophical concepts, autonomy and bias could be represented more accurately in newspaper articles. If autonomy was discussed in newspaper articles, it was usually in relation to science fiction scenarios of fully autonomous robots. It would be more helpful if news articles made clear how the use of AI might impact the autonomy of people on a smaller scale, for example through the deskilling of professionals. Bias was often conflated with prejudice and discrimination in algorithms, even though these topics are usually examples of certain biases in algorithms, it would be better if news media made clear that all algorithms include certain biases, which do not always lead to discrimination.

5.4 Limitations and directions for further research

Even though this research provided new insights through the combination of various theoretical perspectives and methods, there are some limitations to take into account. First of all, the focus on the philosophical concepts that were selected highlighted nuances in how the impact of AI is discussed in the public domain, but it may have drawn attention away from some other important topics. Since the concepts of autonomy, responsibility, fairness, bias, explainability and risk are all relatively abstract they were often not mentioned explicitly. To prevent missing important topics, open codes were used in the expert interviews and media analysis to include related topics and other recurring topics. However, issues that have no clear relation to these philosophical topics, like geopolitical issues related to AI, might require more attention.

Secondly, the amount of participants in the expert interviews was limited to six participants, due to time and budget restrictions. However, since the experts were selected to represent three groups of the quadruple helix, they provided a nuanced and in-depth overview of different perspectives on AI. In this research, focus groups with citizens without expertise in the field of AI were conducted to represent the final group of society in the quadruple helix. A suggestion for further research would be to include a representative of an interest group that focuses on the impact of AI on society in the expert interviews as well, so all four helices are represented by an expert.

Thirdly, the choice was made to conduct a limited large-scale analysis of Dutch newspaper articles, in combination with an in-depth analysis of a smaller sample. The large-scale analysis mainly focused on the division of newspaper articles over time and over different newspapers. In a future study, a data driven analysis with a larger sample of newspaper articles could provide more insight in how often certain themes and topics reoccur in the public debate about AI and its impact. Fourthly, a relatively large amount of the focus group participants had completed their education at a university. A recommendation for a future study would be to include more participants with a lower level of education in focus group discussions about the impact of AI, since this might bring in new perspectives on how AI might impact different people in society.

Finally, some of the studies had to be adapted due to regulations to prevent the spread of the Covid-19 pandemic. Both the expert interviews and the focus groups had to be conducted via online video-conferencing tools. This made it harder to read the body language of the participants and to pick up on nuances in the way they spoke when giving answers. In the focus groups, the online setting of the meeting also made it harder for the participants to spontaneously react to each other. Even though the amount of participants per focus group was reduced in order to facilitate online discussions, in order for everyone to be audible, speaking turns had to be organized more strictly and there was less room for spontaneous interruptions.

5.5 Conclusion

This research contributed to increasing understanding of the science-society relationship, by looking at the role of philosophy and communication in discussions about the impact of artificial intelligence (AI). The combination of the results of an analysis of philosophical literature, expert interviews, a media analysis and focus group interviews showed that the philosophical, expert and public debates about the impact of AI are relatively well-aligned. The philosophical concepts of autonomy, responsibility, fairness, bias, explainability and risk recurred in all three debates, however there were some misalignments in how these concepts were discussed and which concepts received most attention. In order to decrease these misalignments and prevent that important issues are overlooked, news media should add more nuance to their reports about the impact of AI and philosophical literature should focus more on weighing risks and benefits of applying AI in specific contexts, instead of focussing on what risks AI may pose in relation to abstract philosophical concepts.

From a communicative perspective, the comparison of the studies provided new insights in how views about the impact of AI discussed by experts, news media and publics relate to each other. There was a lot of overlap in the content discussed in news media and in the focus groups, pointing towards the reliance of laypeople on news media to receive information about AI. Because of this reliance, it is important that laypeople have a sufficient amount of trust, not only in science, but also in news media, especially in relation to the growing concerns about the current increase of misinformation. In addition, the focus on the philosophical concepts brought out nuances and depth in the analysis of the public debate about AI. This provides new insights about the science-society relationship that can be used to increase understanding of how to deal with emerging technologies in science communication.

Overall, it can be concluded that bringing together insights from philosophy and communication science can help to increase understanding of the science-society relationship. This research showed, that if philosophical, expert and public debates about AI and its impact on society are well-aligned, members of society can get a good understanding of how AI might impact their life. Future research could further increase understanding of the science-society relationship, by comparing the case of AI to discussions about the impact of other emerging technologies in the philosophical, expert and public domain.

References

AI HLEG. (2019a). *A Definition of AI: Main Capabilities and Disciplines*. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

AI HLEG. (2019b). *Ethics Guidelines for Trustworthy AI*. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of Machine Learning Research* (Vol. 81, pp. 1–11).

Brennen, J. S., Howard, P. N., & Nielsen, R. K. (2018). *An Industry-Led Debate: How UK Media Cover Artificial Intelligence*. https://ora.ox.ac.uk/objects/uuid:02126b4c-f4f9-4582-83a0-f8a9d9a65079/download_file?safe_filename=Brennen%2B-%2BUK%2BMedia%2BCoverage%2Bof%2BAI%2BFINAL.pdf&file_format=application%2Fpdf&type_of_work=Report

Brennen, J. S., Schulz, A., Howard, P. N., & Nielsen, R. K. (2019). *Industry, Experts, or Industry Experts? Academic Sourcing in News Coverage of AI*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/industry-experts-or-industry-experts-academic-sourcing-news-coverage-ai

Broek-van den Honingh, N., & de Jonge, J. (2018). Vertrouwen in de wetenschap - Monitor 2018. In *Rathenau Instituut*. https://www.rathenau.nl/nl/wetenschap-cijfers/impact/vertrouwen-de-wetenschap/vertrouwen-de-wetenschap

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

Carayannis, E. G., & Campbell, D. F. J. (2009). "Mode 3" and "Quadruple Helix": Toward a 21st century fractal innovation ecosystem. *International Journal of Technology Management*, *46*(3–4), 201–234. https://doi.org/10.1504/ijtm.2009.023374

Carter, S. M., Rogers, W., Win, K. T., Frazer, H., Richards, B., & Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *Breast*, *49*, 25–32. https://doi.org/10.1016/j.breast.2019.10.001

Chuan, C. H., Tsai, W. H. S., & Cho, S. Y. (2019). Framing artificial intelligence in American newspapers. *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 339–344. https://doi.org/10.1145/3306618.3314285

Cowls, J., & Floridi, L. (2018). *Prolegomena to a White Paper on an Ethical Framework for a Good AI Society*. http://dx.doi.org/10.2139/ssrn.3198732

Cyr, J. (2016). The Pitfalls and Promise of Focus Groups as a Data Collection Method. *Sociological Methods & Research*, *45*(2), 231–259. https://doi.org/10.1177/0049124115570065

D'Souza, R. (2018). *Symbolic AI v/s Non-Symbolic AI, and everything in between?* Medium. https://medium.com/datadriveninvestor/symbolic-ai-v-s-non-symbolic-ai-and-everything-in-between-ffcc2b03bc2e

De Boer, C., & Brennecke, S. (2014). Priming and Framing. In *Media en Publiek: Theorieën over media-impact* (7th ed., pp. 201–210). Boom Lemma uitgevers.

Dijkstra, A. M., & Schuijff, M. (2016). Public opinions about human enhancement can enhance the expert-only debate: A review study. *Public Understanding of Science*,

25(5), 588–602. https://doi.org/10.1177/0963662514566748
Dobbe, R., Dean, S., Gilbert, T., & Kohli, N. (2018). A broader view on bias in automated decision-making: Reflecting on epistemology and dynamics. *ArXiv*.
Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *ITCS 2012 - Innovations in Theoretical Computer Science Conference*, 214–226. https://doi.org/10.1145/2090236.2090255
Ensmenger, N. (2018). The environmental history of computing. *Technology and Culture*, *59*(4), S7–S33. https://doi.org/10.1353/tech.2018.0148
Entman, R. M. (1993). Framing : Toward Clarification of a Fractured Paradigm. *Journal of Communication*, *43*(4), 51–58.
Fiorino, D. J. (1990). Citizen Participation and Environmental Risk : A Survey of Institutional Mechanisms. *Science, Technology, & Human Values*, *15*(2), 226–243.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
Friese, S. (2020). *Measuring Inter-coder Agreement – Why Cohen's Kappa is not a good choice*. ATLAS.Ti. https://atlasti.com/2020/07/12/measuring-inter-coder-agreement-why-cohens-kappa-is-not-a-good-choice/
Guest, G., Namey, E., & McKenna, K. (2017). How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. *Field Methods*, *29*(1), 3–22. https://doi.org/10.1177/1525822X16639015
Hayes, P., van de Poel, I., & Steen, M. (2020). Algorithms and values in justice and security. *AI and Society*, *0123456789*. https://doi.org/10.1007/s00146-019-00932-9
Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, *13*(1), 76. https://doi.org/10.1007/s12178-020-09600-8
Johnson, D. G., & Verdicchio, M. (2017). Reframing AI Discourse. *Minds and Machines*, *27*(4), 575–590. https://doi.org/10.1007/s11023-017-9417-6
Johnson, D. G., & Verdicchio, M. (2018). AI, agency and responsibility: the VW fraud case and beyond. *AI and Society*, *34*(3), 639–647. https://doi.org/10.1007/s00146-017-0781-9
Joseph, M., Kearns, M., Morgenstern, J., & Roth, A. (2016). Fairness in Learning: Classic and contextual bandits. *Advances in Neural Information Processing Systems*, *Nips*, 325–333.
Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information Communication and Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087
Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, *50*(6), 2855–2870. https://doi.org/10.1111/bjet.12868
Liu, Y., Radanovic, G., Dimitrakakis, C., Mandal, D., & Parkes, D. C. (2017). Calibrated fairness in bandits. *Proceedings of FAT-ML*. https://doi.org/10.1145/nnnnnnn.nnnnnnn
Natale, S., & Ballatore, A. (2017). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of*

Research into New Media Technologies, *26*(1), 3–18. https://doi.org/10.1177/1354856517715164
National Academy of Sciences. (2017). *Communicating Science Effectively: A Research Agenda*. https://doi.org/10.17226/23674
O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220
Raab, C. D. (2020). Information privacy, impact assessment, and the place of ethics ∗. *Computer Law and Security Review*, *37*. https://doi.org/10.1016/j.clsr.2020.105404
Reinsborugh, M. (2017). Science fiction and science futures: considering the role of fictions in public engagement and science communication work. *Journal of Science Communication*, *16*(4), 1–8. http://unsettlingscientificstories.co.uk/imagined-futures.
Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C., & Liu, Y. (2020). How do fairness definitions fare? Testing public attitudes towards three algorithmic definitions of fairness in loan allocations. *Artificial Intelligence*, *283*, 103238. https://doi.org/10.1016/j.artint.2020.103238
Schäfer, M. S. (2017). How Changing Media Structures are Affecting Science News Coverage. In K. H. Jamieson, D. M. Kahan, & D. A. Scheufele (Eds.), *The Oxford Handbook of the Science of Science Communication* (pp. 51–59). Oxford University Press.
Scheufele, D. A. (2014). Science communication as political communication. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 13585–13592. https://doi.org/10.1073/pnas.1317516111
Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115
Schothorst, Y., & Verhue, D. (2018). *Nederlanders over Artificiële Intelligentie*. https://www.rijksoverheid.nl/documenten/rapporten/2018/10/31/nederlanders-over-artificiele-intelligentie
Schütz, F., Heidingsfelder, M. L., & Schraudner, M. (2019). Co-shaping the Future in Quadruple Helix Innovation Systems: Uncovering Public Preferences toward Participatory Research and Innovation. *She Ji*, *5*(2), 128–146. https://doi.org/10.1016/j.sheji.2019.04.002
Siune, K., Markus, E., Calloni, M., & Felt, U. (2009). Challenging futures of science in society. *Report of the MASIS Expert …*, 84. https://doi.org/10.2777/467
Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for deep learning in NLP. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, *1*, 3645–3650. https://doi.org/10.18653/v1/p19-1355
Vakkuri, V., & Abrahamsson, P. (2018). The key concepts of ethics of artificial intellligence: A keyword based systematic mapping study. *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1–6. https://blog.growthbot.org/the-ethics-of-artificial-intelligence
Verhue, D., & Mol, P. (2018). *Kunstmatige Intelligentie: Een onderzoek naar de kennis en houding van burgers en ondernemers ten aanzien van Kunstmatige Intelligentie* (Issue November).

https://www.rijksoverheid.nl/documenten/rapporten/2018/11/30/kunstmatige-intelligentie
Weingart, P., & Guenther, L. (2016). Science communication and the issue of trust. *Journal of Science Communication*, *15*(05), 1–11.

Appendix A: Interview protocol

Introduction:

For my master thesis I'm doing research about the expectations that people have about artificial intelligence. I will compare how experts' expectations about machine learning algorithms differ from expectations of the public and explore what role news media play in this case. During this interview I will first ask you about your work and how it involves machine learning. Following this I will ask some questions about your expectations of machine learning algorithms and its possible societal impacts. Finally, I will ask some questions about how you perceive communication in news media about machine learning and artificial intelligence.

As mentioned in the consent form I will record this interview and transcribe it afterwards. In these transcripts you will be pseudonymized and the data from this interview will only be used for this research project in such a way that it can't be traced back to you as a person. If you have any questions about this interview or the research project you can ask them at any time during the interview or afterwards via email. If you would like to withdraw from participating in this research at any time, you can do so by letting me know during the conversation or afterwards via e-mail. In that case your data will be removed from the research and deleted.
Ask if everything is clear.

Topic 1: Job and expertise [3-5 minutes]

1. What is your job?
    a. How does your work involve Machine Learning algorithms?
    b. What does your daily work with machine learning look like?
    c. How did you get involved in this field?

Topic 2: Expectations about Machine Learning [5-10 minutes]

1. What, in your view, is Machine Learning?
    a. How does this relate to artificial intelligence?
    b. How does this relate to other related concepts (like deep learning, data science)?

2. In your view, what is the most impressive possibility with Machine learning at its current state of development?

3. What are your expectations about Machine Learning for the future?
    a. What developments do you expect in the coming few years?
    b. What are your expectations for what will be possible with machine learning technologies the further future (for example in 10, 20, 50 years)?

Topic 3: Societal impact [15-20 minutes]

1. How do you think machine learning technologies might impact society currently?
    a. How would you define societal impact?
    b. What societal impacts are specific to your work or the machine learning applications you work with?
    c. How do you think machine learning technologies might impact…
        i. Different aspects of society: (science, policy/law, industry/economy, social interactions, daily life)?
            - Do you know which aspects are influenced more and which are influenced less?
        ii. Different stakeholders (specifically: scientists, policy makers, industrial stakeholders, citizens, journalists)?
            - Do you know which stakeholders are influenced more and which are influenced less?
    d. How do you think the societal impact of machine learning might change in the near future?

2. How do you think machine learning is perceived by other people?
    a. Do you think your perception of machine leaning and its implications is similar to the perception of other experts in your field?
    b. How do you think other stakeholders perceive machine learning and artificial intelligence? (Scientists, policy makers, industrial stakeholders, citizens)
    c. How do you think these stakeholders assess the societal impact of machine learning?

3. Are there any customs or procedures in your work that draw attention to possible ethical and societal implications of your work?
    a. To what extent are there procedures in place that guide ethical conduct?
    b. If so, can you describe these customs and procedures?
        i. What do they aim for?
        ii. Why were they put in place?
        iii. What do you think of these procedures?

Topic 4: Communication [10-15 minutes]

1. What is your opinion about how machine learning and AI are discussed in news media?
    a. Do you recognize certain themes in news articles about AI?
    b. To what extent is AI discussed in a positive or negative way in news articles
    c. What is your opinion about how the societal impact of AI is discussed in news media?

2. What is your opinion on the public debate about machine learning and AI?
    a. What role do you think communication plays in the public perception of AI?
    b. Do you think the public debate about AI should change? And if so, how?
    c. When would the public debate reflect well what is happening in the field of AI and machine learning?

3. To what extent do you play a role in informing others about machine learning?
    a. Who do you inform about machine learning? (specifically: scientists, policy makers, industrial stakeholders, citizens, journalists)
    b. What role do you think you should play in the public debate about AI?
    c. What do you think you could contribute to the public debate about AI?

4. On what sources do you base your own knowledge about machine learning?
    a. Where do you think others get information about machine learning? (specifically: scientists, policy makers, industrial stakeholders, citizens, journalists)

Conclusion

1. Is there anything you would like to add? (This can be something we haven't discussed yet or something relaed to one of the previous topics we talked about).

Thank you for your participation. *Ask if they have any questions about the research project.*

Appendix B: Codebook expert interviews

Concepts from theoretical framework

- Autonomy
- Bias
- Explainability
- Fairness
- Responsibility
- Privacy

Expertise

- Explanation of work
- Explanation of technology
- Example of technology
- Example of impact

Specified concepts

- Comparison of artificial and human intelligence
- Conscience
- Discrimination/prejudice
- Efficiency
- Humanity
- Human-tech relationship
- Morality
- Power
- Reliability
- Surveillance
- Transparency
- Unspecified ethical/societal implications

Response to / evaluation of technology

- Benefit
- Fear
- Hope
- Risk
- Affordance
- Limitation
- Promise

Recurring themes

- AI hype
- Climate change
- Corona virus
- Dutch Politics

- Geopolitics
- Games
- Education
- Healthcare
- Fake news / disinformation
- Filter bubble
- Law
- Targeted marketing (under example of tech?)

Sources mentioned

- Academia
- Industry
- Politics/governance
- Interest groups/ NGO’s etc.
- Citizens
- Media
- Artists

Appendix C: References newspaper articles

Albers, C. (2019, October 2) Algoritmes tonen ons fundamentele misstanden in de maatschappij. *De Volkskrant,* p. 23

Betlem, R. (2020, August 31) ECB waarschuwt voor dominante positie techgiganten uit VS. *Het Financieele Dagblad,* p. 5

Boon, A. (2020, Januari 18) Juiste diagnose met kunstmatige intelligentie. *Nederlands Dagblad.* p. 18

Bouman, H. (2019, September 13) Prima thuis in de 21ste eeuw. p. 13

Broekhuizen, K. (2019, October 17) 'Ik wil meer vijanden kunnen doden per liter kerosine'. *Het Financieele Dagblad*. p. 10

Broekhuizen, K. (2020, Januari 3) Tech maakt verdeling van werk grootste uitdaging. *Het Financieele Dagblad*. p. 8

Broekhuizen, K. (2020, July 22) Zaak over algoritmes Uber waarschuwing voor andere bedrijven. *Het Financieele Dagblad.* p. 15

Broekhuizen, K. (2020, June 20) Kunstmatige intelligentie is nog altijd vrij dom. *Het Financieele Dagblad.* p. 21

Broekhuizen, K. (2020, June 27) 'Technici laten anderen nadenken over sociale gevolgen'. *Het Financieele Dagblad*. p. 12

Bronzwaer, S. (2019, October 9) Miljarden voor kunstmatige intelligentie. *NRC.NEXT* p. 1

Cath-Speth, C. & Kaltheuner, F. (2020, March 11) EU laat met voorstel kunstmatige intelligentie cruciale kansen liggen. *Het Financieele dagblad*. p. 25

Clahsen, A. (2020, June 27) Dankzij data zit je in de nieuwe Kuip altijd goed. *Het Financieele Dagblad*. p. 6

Cremers, R. (2020, Januari 11) Student geneeskunde moet over robots leren. *NRC Handelsblad*. p. 1

Data rukken op in verzekeringen, zorgen over discriminatie groeien (2020, July 24) *Het Financieele Dagblad.* p. 18

Enghusen, M. (2020, August 13) In high-tech-Israël is art vaker overbodig. *Het Financieele Dagblad.* p. 18

Februari, M. (2019, September 24) Nieuwe technologie? Investeer in wijsheid. *NRC.NEXT*, p. 18.

Funnekotter, B. (2020, Januari 4) Van oermens tot kunstmatige intelligentie. *NRC Handelsblad*. p. 1

Hofman, F. (2020, August 25) Nepvideo Buma blijkt overtuigend. *NRC.Next* p. 2

Holslag, J. (2020, August 29) Deal Huawei mes in de rug van moedige universiteiten. *Het Financieele Dagblad*. p. 30

Hoos, H., Verheij, B.& Van Den Hoven, J. (2020, February 25) Nederland, pak kans met Kunstmatige intelligentie. *De Volkskrant*. p. 23

Kalse, E. (2020, July 10) Hoe Wopke Hoekstra vast bleef houden aan zijn zelfbedachte fonds. *NRC Handelsblad.* p. 1

Kist, R. (2019, October 9) ‘Het baasje is verantwoordelijk, óók voor slimme machines’. *NRC Handelsblad*. p. 4

Loss, L. (2020, August 11) Het gevaar van gezichtsherkenning: meer racisme. *Het Financieele Dagblad.* p. 18

Menselijk gezicht (2019, December 7) *Het Financieele Dagblad*. p. 60

Meulder, M. (2020, July 4) Het algoritme als financiële speurneus. *Het Financieele Dagblad*. p. 6

Nauta. H. (2020, July 8) Datacenters staan al vol met Huawei-apparatuur. *Trouw.* p. 6, 7

Noordermeer, B. (2020, August 29) Etnisch profileren met gezichtsherkenning. *Het Financieele Dagblad*. p. 18

Rootselaar, F. (2020, august 22) ‘We hebben maar één generatie om deze planeet te redden’. *Het Financieele Dagblad*. p. 6

Rotman, R. (2020, May 16) Steken computers de Beatles en Jay-Z naar de kroon? *Het Financieele Dagblad*. p. 18

Schaake, M. (2020, July 10) De lessen van de online burgerrechtenrevolte. *NRC.Next*. p. 6

Schaake, M. (2020, March 7) ‘Het poldermodel is ideaal voor techkwesties’. *Het Financieele Dagblad*. p. 18

Schiffers, M. (2020, February 20) Brussel wil wedloop om kunstmatige intelligentie op ethische wijze winnen. *Het Financieele Dagblad*. p. 8

Schoonen, W. (2020, August 29) Laat Kunstmatige intelligentie haar eigen gang gaan. *Trouw.* p. 14, 15

Software leert drone vliegen. (2019 October 11) *De Telegraaf*, p. 22

Steinbuck, M (2020, June 20) Kunstmatige intelligentie der dingen. *Het Financieele Dagblad.* p. 22

Stinson, C. (2020, June 13) Algoritmes met een duister randje. *Het Financieele Dagblad*. p.20

Tol, N. (2020, August 24) Leeg stadion vol door ‘app-fans’. *De Telegraaf*. p. 9

Van Bemmel, Noël (2020, August 1) Aan het front van de techoorlog. *De Volkskrant*. pp. 25-26

Van Benthem, Jan (2020, Januari 23) Geen autonome dodelijke wapens. *Nederlands Dagblad*. p. 8

Van Lindenburg, H. (2020, August 13) Kunstmatige intelligentie als ‘Het kapitaal’. *Nederlands Dagblad.* p. 12

Van Lonkhuyzen, L. (2019, October 19) Alice lacht niet als je je gebit uitdoet. *NRC Handelsblad* p. 1

Van Lonkuyzen, L (2020, August 8) Siemens breidt uit in kankertherapie. *NRC Handelsblad*. p. 6

Van Noort, W. (2020, March 6) Algoritmes en drones moeten het coronavirus in toom houden. *NRC.Next*. p. 14

Van Sprundel, M. (2020, August 8) De robot wint nog niet (gelukkig). *Trouw.* p. 36

Van Teeffelen, K. (2020, Januari 3) Zo futuristisch is 2020 nog niet. *Trouw*. pp. 2-3

Van Turnhout, M. (2020, July 3) Journalisten maken van de feiten een eigen verhaal. *Trouw.* p. 8, 9

Van Wijnen, J. F. (2019, December 5) De computer die braver is dan de mens. *Het Financieele Dagblad*. p. 20

Van Wijnen, J. F. (2019, October 9) Kabinet vaag over extra budget voor kunstmatige intelligentie. *Het Financieele Dagblad.* p. 9

Van Wijnen, J. F. (2020, August 26) Coalitie kritisch over deal universiteiten met Huawei. *Het Financieele Dagblad*. p. 2

Van Wijnen. J.F. (2020, May 20) AI uit utrecht zoekt juiste corona-artikelen. *Het Financieele Dagblad*. p. 23

Verhagen, L. (2019, December 21) Kunstmatige intelligentie is de mens steeds vaker te slim af. *De Volkskrant*, p. 9.

Verhagen, L. (2020, August 26) 'We mogen van Huawei alles publiceren'. *De Volkskrant*. p. 14

Westerterp, M. (2019, November 13) Games en kunstmatige intelligentie helpen elkaar ontwikkelen. *Nederlands Dagblad*. p. 12

Winkel, R. (2019, September 30) Kunstmatige intelligentie bij werving staat nog aan begin. *Het Financieele Dagblad*. p. 20.

Appendix D: Codebook media analysis

Codes on article level

Newspapers:

- Volkskrant
- NRC Handelsblad and NRC Next
- Het Financieele Dagblad
- Trouw
- Nederlands Dagblad
- De Telegraaf

Type of article:

- News
- Background
- Opinion
- Editorial
- Foreign affairs
- None/unknown

Section:

- Science
- Business
- Technology
- Society / Interest
- Culture
- Special issue
- Front page
- None/Unknown

Frames (from Chuan et al. (2019)):

Impact framing:

- Societal impact
- Group impact
- Personal impact

Issue framing:

- Thematic issue
- Episodic issue

Codes on sentence level

Frames (from Chuan et al. (2019)):

- Risk
- Benefit

Sources mentioned:

Which sources are mentioned? (If there are quotes from people, to what group to they belong?)

- Academia
- Industry
- Politics/governance
- Interest groups/ NGO's etc.
- Citizens
- Media
- Artists

Philosophical concepts:

- Unspecified ethical/societal issues
- Privacy
- Discrimination and prejudice
- Diversity
- Power relations
- Transparency
- Humanity
- Surveillance
- Comparison of artificial and human intelligence (e.g. AI becoming smarter, taking over humans)
- Conscience
- Efficiency
- Human-technology relationship
- Morality (e.g. creation of "moral AI")

From Theoretical framework (focus on more specific codes above first, use these only when explicitly mentioned)

- Autonomy
- Responsibility
- Fairness
- Explainability
- Bias

Recurring topics:
- Corona virus
- Geopolitics
- games
- healthcare
- climate change
- fake news/disinformation
- law/regulation

Reactions to technologies:
- hope
- fear
- reliability (how reliable and accurate is the technology)
- affordance (new possibilities caused by technology)
- limitation
- promise (expectation of what may become possible)

Appendix E: Focus group protocol

1. Introduction (5 minutes)

- Welcome and thanks for participating.
- Short explanation of research and goal.
- Explanation of focus group: Discussion, use raise hand button or chat to react. No right or wrong answers, goal is to get different perspectives and opinions.
- Introduction round: Mention name, age, job and highest level of education.

2. Introductory questions (10 minutes)

- What do you think about when you hear the term "artificial intelligence"?
- How would you describe artificial intelligence?
- What applications of artificial intelligence do you know about?
- According to you, what is the most impressive application of AI that's currently possible?

- Where have you heard about artificial intelligence?
- What media do you usually use to get informed about the news in general?
- Do you occasionally read news articles about AI? If so, what do you think of how it is discussed? (e.g. positively or negatively, do you notice recurring themes?)
- How much trust do you have in the following stakeholders? (categorize from most to least trust: scientists, government, companies, traditional media and social media)

3. Specific questions (10 minutes)

Explanation of artificial intelligence (in Dutch):

De EU en Nederlandse overheid gebruiken de volgende definitie van kunstmatige intelligentie: "Kunstmatige intelligentie verwijst naar systemen die intelligent verdrag vertonen door hun omgeving te analyseren en met een zekere mate van zelfstandigheid actie ondernemen om specifieke doelen te bereiken." Kunstmatige intelligentie is dus een overkoepelend begrip voor verschillende technologieën en technieken, die vaak samen gebruikt worden.

Als we het over kunstmatige intelligentie hebben gaat het vaak over algoritmes, dat zijn bepaalde formules waardoor computers zelfstandig bepaalde taken uit kunnen voeren. Je kunt recepten ook zien als een soort simpel algoritme, het geeft regels waarmee je een bepaald gerecht kunt maken (bijvoorbeeld ALS de ui glazig is DAN moet je de rest van de groente toevoegen).

Tegenwoordig wordt er steeds meer gebruik gemaakt van zelflerende algoritmes (machine learning). Dit is een speciaal soort algoritme, waarbij de regels niet één voor één door een programmeur worden uitgeschreven, maar waarbij een computer zelf patronen leert herkennen in data die door een programmeur wordt ingevoerd. Een bekend voorbeeld hiervan is dat een algoritme door pixels in een digitale afbeelding te analyseren kan leren of er een hond of een kat op een foto staat. De computer gokt eerst willekeurig of iets een hond of kat is, de programmeur geeft feedback of dat klopt en daardoor leert het algoritme langzaam steeds beter honden en katten te herkennen.

- Is this explanation clear or do you still have questions before we continue?
- How does this explanation fit with what you thought about AI before?
- What benefits of AI can you think of?

- What risks of AI can you think of?
- What impact do you think AI has on society? (e.g. on your own life or on specific groups of people)

4. Scenarios (30 minutes)

Discuss two out of four fragments from news articles about various applications of AI and their impact.

Use the following questions for each scenario:

- What is your first reaction to this news article?
- What is your opinion of this application of AI? (Is this an appropriate context to use AI? Is AI used well in this case?)
- Are there any benefits of using AI in this case? Are there any risks? What are they?
- Do the benefits outweigh the risk?
- If applicable: Would you like to use this application of AI?
- Do you think the use of AI in this case and similar situations affects your daily life?
- Do you think this news fragment accurately describes the AI application and context it discusses?
- How much trust do you have in the stakeholders mentioned in this fragment?

News fragment 1: Discussion of the use of AI in facial recognition software in surveillance and how it might lead to racism.

News fragment 2: Researchers from the University of Amsterdam created a deepfake video of a Dutch politician (Sybrand Buma) and discovered that it influenced people's opinion about him, especially if they were likely to vote for his party.

News fragment 3: Explanation of the use of AI (image recognition algorithms) to help doctors detect prostate cancer and assess how aggressive it is.

News fragment 4: An interview with philosopher Luciano Floridi about how AI can help to solve important problems, especially anthropogenic climate change.

5. Conclusion (5 minutes)

- Has your view of AI and how it might impact society changed after the explanation and the examples we discussed?
- After learning more about artificial intelligence, what developments do you expect in this area in the next 5 years?
- Do you think the application of artificial intelligence systems might impact your own life? If so, how?

- Is there anything we haven't discussed yet that you would like to add?

Thank you for your participation!

News fragment 1:

Een algoritme in software voor gezichtsherkenning produceert op basis van een input een output. Het verwerkt een beeld en beslist of daarin een menselijk gezicht is te zien of niet. Om te leren hoe een gezicht eruitziet, wordt het algoritme getraind met voorbeelden van gezichten. Op een bepaald moment herkent het in de beelden terugkerende patronen, kenmerken en structuren.

Als de verzameling trainingsbeelden niet divers is, zegt computerwetenschapper Joy Buolamwini, dan zijn gezichten die te sterk afwijken van de opgestelde norm voor het algoritme moeilijker te herkennen. Gezichtsherkenningssoftware die alleen met beelden van witte mensen getraind wordt, ziet geen mensen van kleur.

De algoritmen van Amazon, Apple, Facebook en andere concerns zijn black boxes 'Als we met mensen praten, kunnen we op verschillende manieren nagaan hoe hun beslissingen tot stand komen en of die beslissingen discriminerend zijn', zegt Sarah Chander, van de ngo European Digital Rights in Brussel.

Wie black boxes maakt van algoritmen, verhindert dat beslissingsprocessen openbaar zijn en zegt tevens dat slechts een heel kleine groep kan begrijpen hoe applicaties met kunstmatige intelligentie werken. Dat leidt uiteindelijk tot een steeds grotere machtsongelijkheid tussen de mensen die kunstmatige intelligentie ontwikkelen en toepassen, en degenen die overgeleverd zijn aan hun zogenaamd objectieve beslissingen.

'Ook wanneer zulke systemen data niet vooringenomen behandelen, produceren ze dus misschien toch een racistische output', zegt Chander. 'Want algoritmen reproduceren het racisme niet alleen, ze versterken het ook. En ze hebben het potentieel om dat op veel grotere schaal te doen dan mensen.'

News fragment 2:

Onderzoeker politieke communicatie Tom Dobber (Universiteit van Amsterdam) verzamelde uren aan beeldmateriaal van toenmalig CDA-leider Sybrand Buma. Met kunstmatige intelligentie trainde hij een algoritme, dat van de oude beelden leerde over Buma's stem en gezichtsbewegingen. De software manipuleerde vervolgens een bestaande video van de politicus, door lipbewegingen en audio uit andere beelden in het betreffende filmpje te plakken. In de vijf seconden durende nepvideo (deepfake) maakt Buma een woordgrap over de kruisiging van Jezus.

,,Het was een realistische video, al zag ik dat de lipbewegingen soms niet helemaal goed gingen", zegt Dobber. Maar van de 140 participanten die de video zagen, vermoedden er slechts acht manipulatie.

Bovendien beïnvloedde de video hun mening over Buma in negatieve zin. Een even grote groep zag alleen de oorspronkelijke video. Alle participanten vulden daarna een vragenlijst in, onder meer over Buma's betrouwbaarheid en vriendelijkheid. De deepfake-groep scoorde een gemiddelde van 4,31 uit een totaal van zeven, ruim 0,3 lager dan de controlegroep.

Nog groter was het verschil onder de mensen die eerder op het CDA hadden gestemd. In de controlegroep kreeg de politicus van hen een gemiddelde score van 5,43, tegenover 4,72 van de Bumafans uit de deepfake-groep.

,,Het verschil zal waarschijnlijk groter zijn wanneer zo'n filmpje langer duurt, of de uitspraken extremer zijn", zegt Dobber. Zeker wanneer de makers een video goed timen en met slechte bedoelingen verspreiden. Stel je voor dat zo'n video grootschalig wordt uitgesmeerd vlak voor verkiezingen. De schade is mogelijk nog groter wanneer zo'n video specifiek aan een kleine

groep gericht is. ,,De kans dat de video dan opvalt bij de media en dus gecorrigeerd wordt, is kleiner."

News fragment 3:

Onderzoekers van de Radboud Universiteit in Nijmegen (waaronder Geert Litjens) hebben een systeem ontwikkeld dat de agressiviteit van prostaatkanker beter kan inschatten dan de meeste pathologen.

Voor het stellen van de diagnose prostaatkanker is microscopische analyse van stukjes weefsel belangrijk. De patholoog kan daarin tumorcellen van gezonde cellen onderscheiden, legt Litjens uit. 'Juist bij prostaatkanker is het heel belangrijk om in te schatten of die tumorcellen dusdanig agressief zijn dat die patiënt eraan gaat overlijden.' Een overschatting van de agressiviteit zou kunnen betekenen dat een tachtigjarige een zware operatie en chemotherapie ondergaat, terwijl die persoon nooit aan deze tumor zou overlijden.

Tien jaar geleden instrueerde Litjens de computer nog waar die bij analyse van de foto's precies naar moest kijken. 'Wij programmeerden echt op deze manier: de cellen moeten zo groot zijn en die vorm hebben. Nu zeggen we: hier zijn de plaatjes, dit is de uitkomst. Ga zelf maar leren wat relevant is om van de plaatjes tot die uitkomst te komen.'

Om nog enigszins te kunnen traceren waarop het systeem zijn conclusies baseert, is de analyse opgesplitst in de verschillende stappen die een patholoog neemt bij beoordeling van het weefsel, legt Litjens uit. 'Je splitst het dus op en de patholoog kan op elk stukje van dat proces zien wat de computer doet en daarop ingrijpen. Als het systeem bijvoorbeeld de verkeerde cellen herkent, hoef je die agressiviteitsbeoordeling ook niet te vertrouwen.'

Het AI-algoritme wordt nog niet in de praktijk gebruikt. Bovendien kunnen ze voorlopig alleen een medisch specialist ondersteunen, niet vervangen, stelt Litjens. 'De systemen zijn specifiek toegerust om de agressiviteit van prostaatkanker te bepalen. Maar een patholoog kan nog een heleboel andere dingen in zo'n weefselplaatje zien waar het systeem niet op getraind is en niks mee doet. Een arts moet daarom altijd een final check doen.'

News fragment 4:

Volgens filosoof Luciano Floridi hebben we kunstmatige intelligentie (AI) nodig in de strijd tegen klimaatverandering. Floridi wordt beschouwd als de founding father van de informatie-ethiek, en is een belangrijk adviseur van de Europese Commissie over informatietechnologie.

Floridi: 'Als het over de ethiek van AI gaat, kun je het hebben over misbruik, overmatig gebruik, maar ook over te weinig gebruik. We zijn geneigd dat laatste punt te vergeten, terwijl de gemiste kansen, de opportuniteitskosten, nu al enorm zijn.' Neem de medische sector, we deinzen ervoor terug om daarvoor te investeren in digitale technologie. Terwijl we door zo'n investering het menselijk lijden kunnen verminderen, beter aan preventie kunnen doen en ook doden kunnen voorkomen.

Op het gebied van klimaatverandering kunnen we dankzij machine learning ons stroomverbruik enorm terugdringen door de efficiëntie van het gebruik te vergroten. Doordat AI zelf ook stroom gebruikt, is er wel altijd een compromis tussen de hoeveelheid energie die de digitale technologie zelf gebruikt en de vermindering van het energiegebruik door ons. Maar de uitkomst daarvan is duidelijk: gebruik van die technologie kan netto een energiebesparing opleveren.

Dat soort AI kunnen we ook gebruiken om het verkeer te coördineren en files te vermijden. Zo dring je het brandstofgebruik terug, en het is ook nog eens aangenamer voor mensen om niet in de file te zitten.

Een belangrijk punt, zegt Floridi, is dat we die kunstmatige intelligentie ook kunnen gebruiken om hernieuwbare energie economisch rendabel te maken. 'Om de energie van zon, wind en golven zo goed mogelijk te gebruiken, en daarmee rendabel te maken, moet je rekening houden met kleine verschillen. Wanneer is er bijvoorbeeld de meeste wind? Voortdurend die kleine verschillen monitoren, dat is precies wat de computer het beste kan.

Appendix F: Codebook focus groups

Concepts from literature review

- Autonomy
- Bias
- Explainability
- Fairness
- Responsibility
- Privacy
- Trust

Specified concepts

- Comparison artificial and human intelligence
- Conscience
- Discrimination/prejudice
- Efficiency
- Humanity
- Human-tech relationship
- Morality
- Power
- Reliability
- Surveillance
- Transparency
- Unspecified ethical/societal implications

Response to / evaluation of technology

- Explanation of technology
- Example of technology
- Example of impact
- Benefit
- Fear
- Hope
- Risk
- Affordance
- Limitation
- Promise

Recurring themes

- AI hype
- Climate change
- Corona virus
- Geopolitics
- Games

- Education
- Healthcare
- Fake news / disinformation
- Filter bubble
- Law
- Targeted content

Sources mentioned

- Academia
- Industry
- Politics/governance
- Interest groups/ NGO’s
- Citizens
- Media
- Artists